

<https://archined.ined.fr>

# **Theory and applications of backward probabilities and prevalences in cross-longitudinal surveys**

**Nicolas Brouard**

**Version**

Accès restreint

## **POUR CITER CETTE VERSION / TO CITE THIS VERSION**

Nicolas Brouard, 2019, "Theory and applications of backward probabilities and prevalences in cross-longitudinal surveys". Dans : Arni S. R. Srinivasa Rao (Éd.) et C.R. Rao (Éd.), Handbook of Statistics. Integrated population biology and modeling, Part B, Elsevier, p. 435-486.

<https://doi.org/10.1016/bs.host.2018.11.009>

**Disponible sur / Available at:**

<http://hdl.handle.net/20.500.12204/AWltfxMuXMQCvuZmm9UI>

# Theory and Applications of Backward Probabilities and Prevalences in Cross-Longitudinal Surveys

Nicolas Brouard<sup>a,1,\*</sup>

<sup>a</sup>*National institute for demographic studies (INED), France*

---

## Abstract

In this chapter we introduce the backward probability and the backward prevalence. Both measures are of minor interest compared to forward probability and forward prevalence; however, they can bring a better understanding of the past population dynamic. Moreover, we show that through the calculation of backward probabilities, one can reconstruct prevalences at any age for older generations. Here, the demographic interest lies in the comparison of the three prevalences: cross-sectional, forward and backward. In order to accomplish our task, we first review theories of Markov chains with *i*) an age-independent transition matrix and, *ii*) when transitions vary with age. The second theory leads to an interesting property called “weak ergodicity” that allows us to predict future prevalences for younger generations. It is important to mention here that backward probability was rapidly defined in 1980 (Brouard, 1980) using longitudinal information of French women’s participation in economic activity between 1977 and 1978, and most results presented in the publication of 1980 are reviewed for this study.

This chapter also shows that in a stationary multi-state population, cross-sectional, forward and backward prevalences are identical at each age.

If they are not, as in the case of the economic Q14 activity of French women which changed after the 1968s revolution (women’s liberation, contraception, and abortion laws), our approach enables a clearer, faster, and synthetic analysis of these changes without the need to wait 20 or 30 years until these women leave the labor market.

If they are not, as in the case of the economic activity of French women which changed after the 1968’s revolution (women’s liberation, contraception and abortion laws), our approach enables a clearer, faster and synthetic analysis of these changes without the need to wait 20 or 30 years until these women leave the labor market.

Then, we review demographic tools, now widely used in mortality analysis that compare ‘cross-sectional prevalence of survival’ and period mortality table. We extend them to multi-states methods, particularly to methods developed in the mid-90s to estimate disability-free life expectancies.

---

\*<https://doi.org/10.1016/bs.host.2018.11.00>

But in current cross-sectional surveys, as in the case of health surveys, people do not respond to the survey at every wave, so the health status may be unknown or known with exposure times unequal between individuals, making multinomial logistic regression useless. A more complex model like the Interpolated Markov chain Model must be applied. Since the software IMaCh has been developed in the late 90's, the use of the Interpolated Markov chain Model greatly increased among health and disability researchers and more accurate estimates of healthy life expectancy can be calculated.

It is also important to mention here that the latest experimental version 0.99 of IMaCh integrates the calculation of the backward prevalences with its confidence intervals. In this chapter, we apply this latest IMaCh version to show the relevance of backward and forward prevalences to understand the aging of societies resorting recent data from cross-longitudinal surveys in Italy and in the United States.

*Keywords:* demography, statistics, multistate, mortality, labor force, cross-longitudinal survey, life expectancy, markov chain, ergodicity, 2000 MSC: 00A71, 47A35

---

## Summary

<b>1</b>	<b>Backward probabilities estimated from chained labor force surveys</b>	<b>4</b>
1.1	Probability or forward probability . . . . .	4
1.2	Backward probability . . . . .	7
1.3	Markov chains and strong ergodicity . . . . .	9
1.3.1	Two states: activity and inactivity by age . . . . .	13
1.3.2	Two states, retroprojection . . . . .	14
1.3.3	Two states, backprobability . . . . .	15
1.4	Weak ergodicity . . . . .	18
1.5	Backward prevalence of a specific cohort . . . . .	21
1.6	Forward prevalence of a specific cohort . . . . .	23
<b>2</b>	<b>Backward probabilities with transient states and an absorbing state</b>	<b>23</b>
2.1	Chaining forward for a specific cohort . . . . .	26
2.2	Chaining backward for a specific cohort . . . . .	31
2.2.1	Stationary population, forward and backward prevalences	33
2.3	Some estimations of backward prevalences . . . . .	43
2.3.1	Did the 2008 economic crisis in Italy impact the Health Expectancies? . . . . .	43
2.3.2	Health and Retirement Study: changes in living arrangements	44
2.4	Limitations concerning the estimation of forward and backward prevalences . . . . .	45
2.5	Perspectives concerning the estimation of backward prevalences .	48
<b>3</b>	<b>Conclusion</b>	<b>49</b>

<b>4</b>	<b>Acknowledgements</b>	<b>50</b>
<b>5</b>	<b>References</b>	<b>50</b>

A backward probability can be defined as the probability to act today while being conditioned by an event which will occur in the future.

A classical probability or “forward” probability is, for example, the probability of having a child two years after marriage; or more generally since the time elapsed since marriage, but a demographer might be interested in studying fertility according to the time remaining before the divorce, because a potential separation could explain a lower fertility. A backward probability can also be considered as a measure of indirect variables and latent variables that are unknown in the present but nevertheless influence the future.

Forward probabilities in Demography, Economics and Statistics are common and powerful. Backward probabilities are clearly less relevant because literature on this concept seems to be rare or even non-existent. But are they completely useless? We will explore their power in a better understanding of our past history using an already published example (Brouard, 1980) as well as to propose a new software, IMaCh 0.99 (2018) <http://euroves.ined.fr/imach> which estimates backward probabilities (as well as their companion the backward prevalences) from adequate data. These data are provided by a particular type of statistical survey that are becoming more common nowadays and are now called longitudinal cross-sectional surveys (Lièvre et al., 2003).

In simple longitudinal surveys, such as labor force surveys, where mortality before retirement age is negligible, backward probabilities are as natural as forward probabilities, but they are different and can be combined to produce backward prevalence. Forward probabilities can be applied to a fictitious young cohort to produce forward age-specific prevalences. Then, they can be compared with current cross-sectional prevalences and even with the forward prevalences of younger cohorts.

If the concept of forward probability and forward prevalences is well-known in the case of a period life table, the concept of backward probabilities and backward prevalences makes sense only in multi-state models.

## 1. Backward probabilities estimated from chained labor force surveys

Let us take the example of a labor force survey where women had their first interview in 1977 describing their economic activity status and interviewed again one year later in 1978 to measure changes. From this follow-up, age-specific input/output matrices can be constructed similar to Table 1 describing the repartition of the  $N_{..}$  women into  $N_{1.}$  active and  $N_{2.}$  inactive women at age  $x$  in 1977 as well as the various flows which occurred in one year up to age  $x + 1$  in 1978.

### 1.1. Probability or forward probability

From this table, we can calculate the age-specific probability (also called forward probability) of leaving the labor force in one year by  $\hat{c}_x = \frac{N_{12}}{N_{1.}}$ , as well as the probability of entering or re-entering the labor force by  $\hat{a}_x = \frac{N_{21}}{N_{2.}}$ .

Table 1: MATRIX OF INPUT/OUTPUT BY AGE: TWO LINKED LABOR SURVEYS.

Age $x + 1$ 1978	ACTIVE	INACTIVE	
Age $x$ 1977			
ACTIVES	$N_{11}$	$N_{12}$	$N_{1.}$
INACTIVES	$N_{21}$	$N_{22}$	$N_{2.}$
	$N_{.1}$	$N_{.2}$	$N_{..}$

We can construct age-specific (forward) transition matrices  $P_x$

$$P_x = \begin{pmatrix} 1 - \hat{c}_x & \hat{c}_x \\ \hat{a}_x & 1 - \hat{a}_x \end{pmatrix}$$

and by right multiplying them, we can calculate the probability of being out of the labor force after  $n$  years for someone in the labor force at age  $x$  as well as the probability of being in the labor force after  $n$  years for someone outside the labor force at age  $x$ , etc.

A step by step description can be done in order to better understand the figures and the ergodic properties of multiplying stochastic matrices:

If the probability for a woman inactive at age 30 in 1977 to be active at age 31 in 1978 is 12%;

- if the probability for a woman inactive at age 31 in 1977 to be active at age 32 in 1978 is 12.04%;

- if the probability for a woman active at age 30 in 1977 to be still active at age 31 in 1978 is 94%;

- if the probability for a woman active at age 31 in 1977 to be still active at age 32 in 1978 is 95%;

then we can write, using the Markov assumption that the probabilities depend only from the current state, that the probability for a woman active at age 30 to be out of the labor at 32 is

$${}_2P_{30} = P_{30} P_{31}$$

$$\begin{pmatrix} 90\% & 10\% \\ 22\% & 78\% \end{pmatrix} = \begin{pmatrix} 94\% & 6\% \\ 12\% & 88\% \end{pmatrix} \begin{pmatrix} 95\% & 5\% \\ 12.04\% & 87.96\% \end{pmatrix}$$

And for a projection in  $k$  years ahead:

$${}_kP_x = P_x P_{x+1} \cdots P_{x+k-1}$$

The product matrix benefits, when  $k$  is increasing, of weak ergodicity properties which can be visually appreciated in Fig. 1, providing a stable curve, which is also the limit of the projections (see Fig. 2).

It has been called the “period” activity curve, because it is computed only from age-specific incidences estimated during the 1977-1978 period of these two first

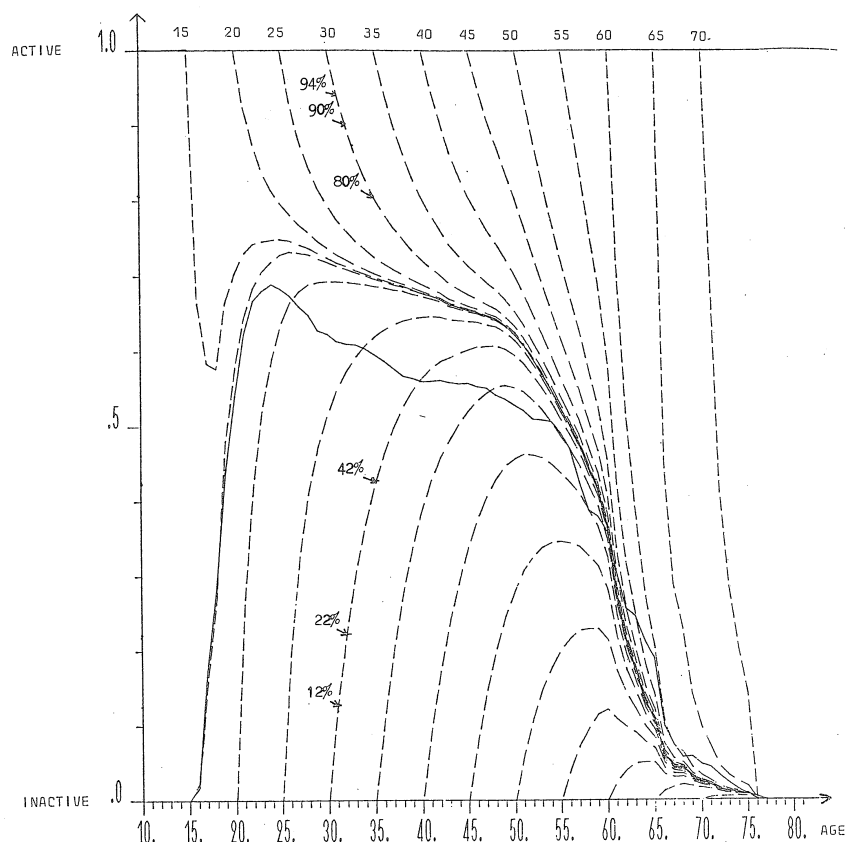


Figure 1: Probability for a woman inactive at age 30 in 1977 to be active at age 31 (12%) in 1978, 32 (22%) in 1979 as well as (on top), probability for a woman active at age 30 to be still active at 31 (94%), 32 (90%) etc. Source: (Brouard, 1980)

chained labor force surveys (Brouard, 1980). Now, it will be also called the forward prevalence of activity.

Comparing the cross-sectional curves of activity by age (in 1977 and 1978) with projected curves and more interestingly with the ultimate stable curve, we can draw some interesting sociological conclusions concerning this period in France.

Because of the development of massive contraception in the mid-1970s (contraception law in 1967, abortion law in 1975), women changed their employment and maternity behavior. It has become easier to postpone the formation of a family or to supplement one's family while retaining a job and enter or re-enter the labor market even if the children were already born.

The period (forward) prevalence is induced by the probabilities  $a_x$  and  $c_x$ , which are represented on the graph 3.

The cross-sectional prevalence observed in 1977 or 1978 is not stable. It is analogous to an age pyramid from a census which differs from the stable population induced by the age specific fertility and mortality rates.

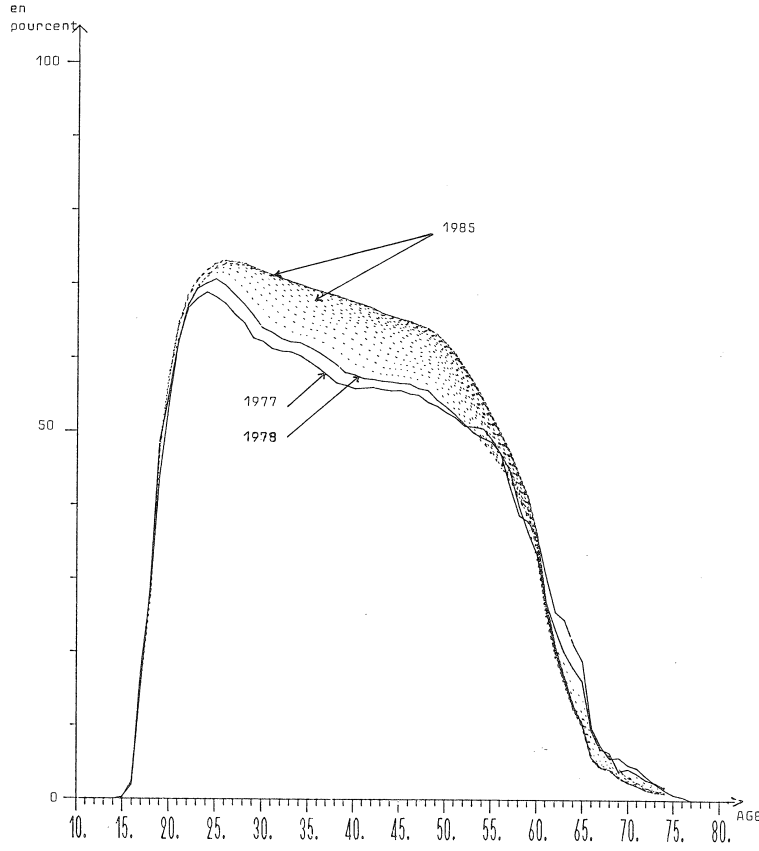


Figure 2: Observed activity ratios by age in 1977 and 1978 and projected in the future until convergence is reached. Projections after year 1985 can be distinguished from the limit only between ages 40 and 50. Source: (Brouard, 1980)

### 1.2. Backward probability

According to Table 1 we can compute a different probability  $b^{12} = \frac{N_{12}}{N_2}$  defined as the probability to be inactive at age  $x$  knowing that we will be active at age  $x + 1$ . It can be called a *backward* probability, as well as for  $b^{21} = \frac{N_{21}}{N_1}$  which is the probability to be active at age  $x$  knowing that will be inactive at age  $x + 1$ .

Let us remark that extending the classical notation in demography for a probability of death  ${}_h q_x$ , the full notation should be  ${}_1 p_x^{ij}$  and  ${}_{-1} b_{x+1}^{ij}$ . We will keep the left index notation only for matrices.

Then we can make backward projections

$$B_{x+1} = \begin{pmatrix} 1 - b^{21} & b^{12} \\ b^{21} & 1 - b^{12} \end{pmatrix}$$

$${}_{-n} B'_{x+1} = B'_{x+1} B'_x \cdots B'_{x-n}$$

and because of the weak ergodicity property of stochastic matrices again we get convergence in the past in Fig. 4.



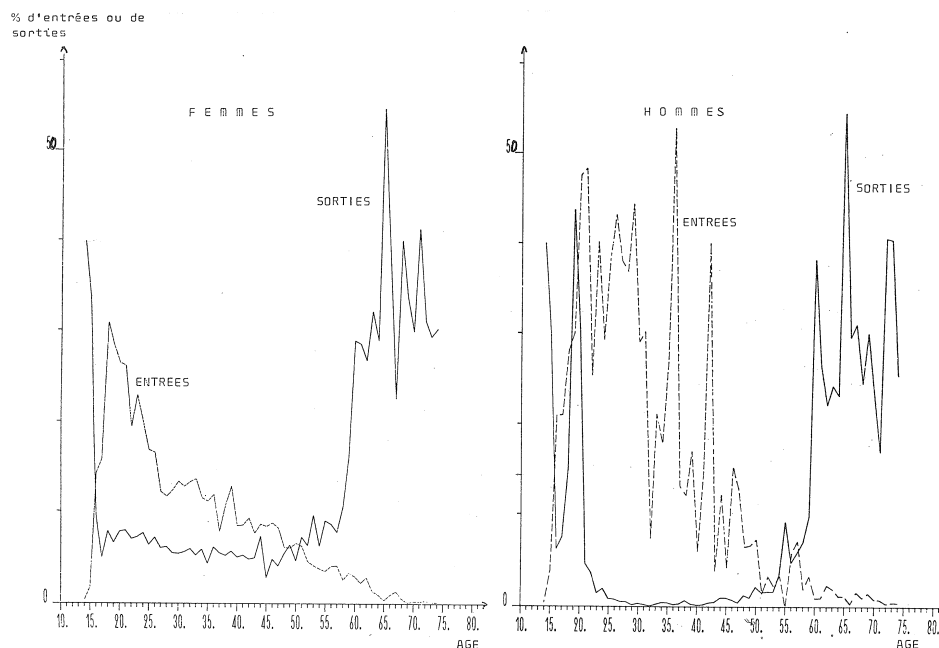


Figure 3: Probability of entering  $a_x$  or leaving  $c_c$  the labor force in France (1977-1978) by sex. Source: (Brouard, 1980)

But what is very interesting in this backward prevalence is that it corresponded to the age-specific activity ratios that prevailed in France before 1968, with the well-known "M" shape: the women were at work young but left their jobs to raise their families and some of them did return to the labor force when the children were already grownup.

We can appreciate the quality of both projections on Fig. 5 by comparing with the activity ratios observed during the censuses of 1968, 1982 and 1990 in France.

We remark that the forward prevalence better fits the prevalence of 1982 at younger ages and that of 1990 at older ages: we can see on Fig 2 that the horizon of 1982 is only 4 years after 1978 and that the horizon of 1990 is 5 years after 1985 and can't be distinguished from the limit.

If the backward prevalence reveals the "M" shape and if this backward projection is correct at older ages proving that the retirement age was higher in 1968 than in 1977-78, the hollow of the wave is not low enough and reflecting also the fact that the limit of the backward projection can't be estimated earlier than in 1972 which is close to stable backward prevalence as shown in the Fig. 6.

From a demographic point of view, the analogy with life table analysis is simple. A life table is nothing more than a projection of cohort survival according to the age-specific mortality rates observed over a period of time. And projections applied to complete cohorts of newborns are less reliable than those applied to middle-aged cohorts.

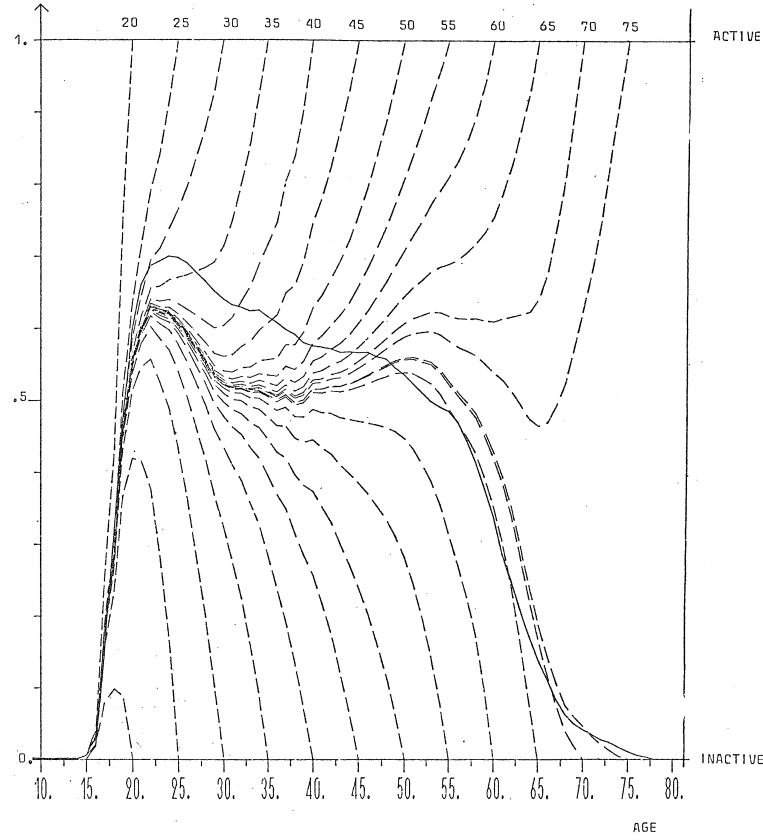


Figure 4: Backward probability to be active at  $x - h$  in a cohort, active or inactive at age  $x$ . Source: (Brouard, 1980)

After this intuitive presentation of the forward and the backward probabilities describing their ability to predict changes in populations by amplifying apparently minor age-related changes, we return to theoretical parts. We will also discuss the divergence observed in Fig. 6 with the retroprojections into the future.

### 1.3. Markov chains and strong ergodicity

In this section, we will suppose that the transition matrix  $P_x$  is independent of age, which means that  $c_x = c$  and  $a_x = a$ . We will review the properties of a simple Markov chain, the convergence into the future, but also the divergence into the past (retroprojection). Then, we will introduce the back probabilities which will induce a different Markov chain, which will converge into the past and subsequently diverge into the future.

Let us remember some properties of a Markov chain with  $n$  states. The transition matrix  $P$  of a Markov chain is a stochastic matrix which means that the sum of the coefficients of each row is one ( $\sum_{j=1,n} p_{ij} = 1, \forall i$ ). Let us suppose for simplicity that the  $n$  roots, real or complex,  $\lambda_1, \lambda_2, \dots, \lambda_n$  of  $\det(P - \lambda I)$  are distinct

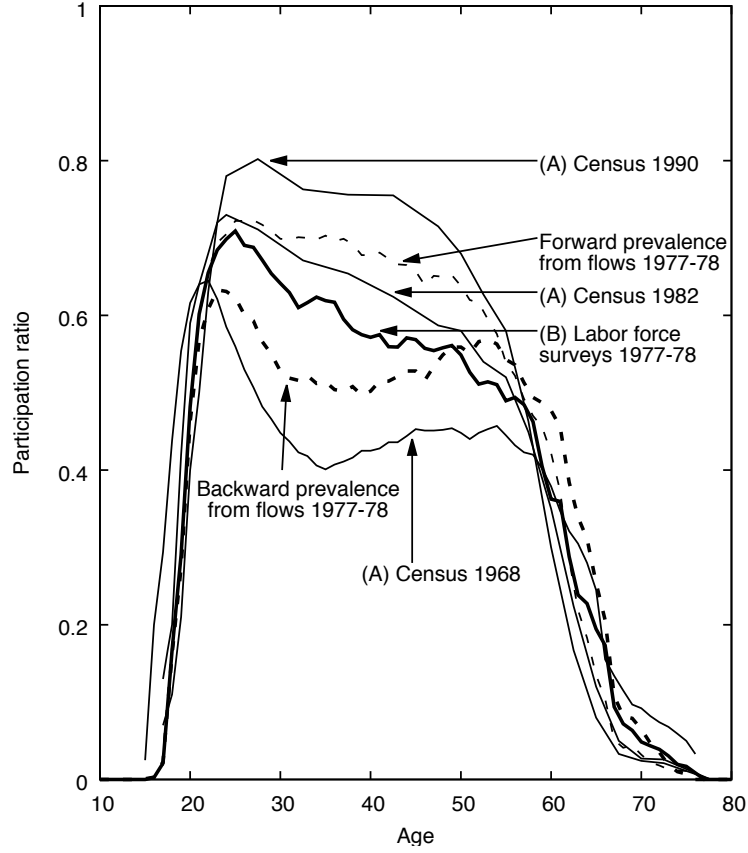


Figure 5: Female labor participation ratios by age from (a) censuses in 1968, 1982, 1990 (b) from the labor force surveys in 1977-78 (raw data). Comparison with the forward prevalence and backward prevalence computed from the chained labor forces in 1977 and 1978.

and not equal to zero. Then the space engendered by the  $n$  eigen vectors is also of dimension  $n$ . The matrix  $P$  can be diagonalized in the space of the  $n$  latent vectors.

Let  $U$  be the matrix of the  $n$  right eigen vectors in column  $(U_i, i = 1, n)$ ,  $V^\top$  the transposed matrix of the left eigen vectors in column  $(V_i^\top, i = 1, n)$  and  $\Lambda$  the diagonal matrix of the  $n$  eigen values ordered by decreasing magnitude  $|\lambda_1| > |\lambda_2| > \dots > |\lambda_n|$ . Then, we can write:

$$\begin{aligned}
 PU_i &= \lambda_i U_i \\
 V_j^\top P &= \lambda_j V_j^\top \\
 V_j^\top PU_i &= \lambda_j V_j^\top U_i = V_j^\top \lambda_i U_i \\
 \text{or } (\lambda_j - \lambda_i) V_j^\top U_i &= 0 \\
 \text{and } V_j^\top U_i &= \begin{cases} 0 & \text{if } i \neq j \\ k_i & \text{if } j = i \end{cases}
 \end{aligned}$$

It is easy to normalize the eigen vectors such that  $k_i = 1$  and  $V_j^\top U_i = \delta_{ij}$  (where

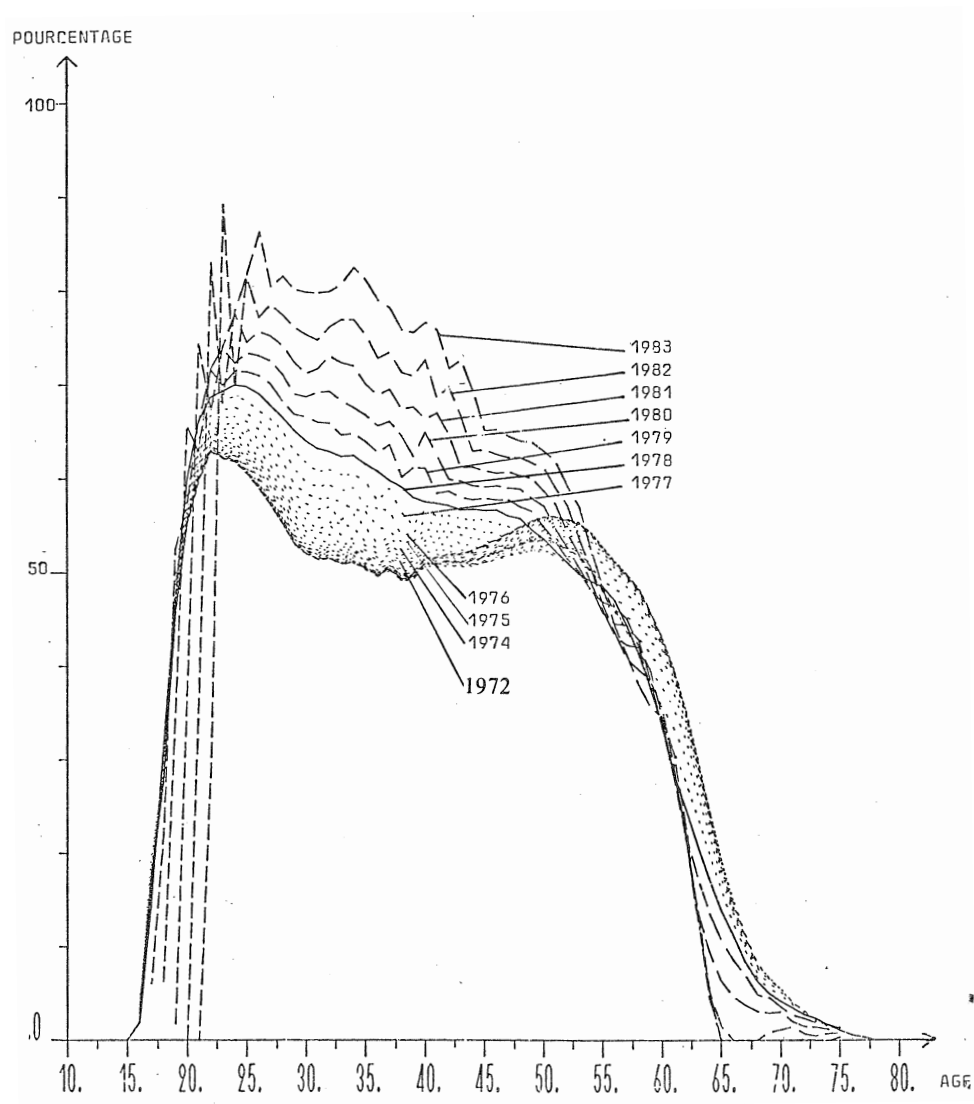


Figure 6: Backward projections of female participation rates (France 1977-78). Backward convergence to the situation before 1968. Years before 1972 are hard to be distinguished from the limit. Retroprojections into the future are divergent. Source: (Brouard, 1980)

$\delta_{ij}$  is equal to 1 if  $i = j$  and 0 if  $i \neq j$ ),  $V^\top U = I$ ,  $U^\top V = I$  so that  $U^{-1} = V^\top$ .

Then we can write

$$\begin{aligned} PU &= U\Lambda \\ V^\top P &= \Lambda V^\top \\ \Lambda &= \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix} \\ V^\top PU &= \Lambda V^\top U = \Lambda \end{aligned}$$

and by right multiplying first line by  $V^\top$

$$PUV^\top = P = U\Lambda V^\top = \sum_{i=1,n} \lambda_i U_i V_i^\top = U_1 V_1^\top + \sum_{i=2,n} \lambda_i U_i V_i^\top$$

Also,

$$P^2 = U\Lambda V^\top U\Lambda V^\top = U\Lambda^2 V^\top \quad (1)$$

$$P^k = U\Lambda^k V^\top = \sum_{i=1,n} \lambda_i^k U_i V_i^\top \quad (2)$$

Because of the stochasticity of the matrix  $P$ ,  $\lambda = 1$  is a latent root and  $e^\top = (1, 1, \dots, 1)^\top$  is a (right) latent vector,  $Pe = e$ . Let  $V_1^\top$  be the corresponding left latent vector with latent value to 1, it is also a latent vector for any matrix  $P^k$  and is the stationary distribution of probabilities

$$\begin{aligned} V_1^\top P &= V_1^\top = V_1^\top P^k \quad \forall k \\ V_{1h} &= \sum_j V_{1j} P_{jh} \quad \forall h \end{aligned} \quad (3)$$

If a left latent vector  $W$  of  $P$  is associated to a latent value  $\lambda$ , it is also the left latent vector of the iterated matrix  $P^k$  with latent value  $\lambda^k$

$$\lambda^k W_h = \sum_j W_j P_{jh}^k \quad \forall h \text{ and } k \quad (4)$$

$$|\lambda^k| |W_h| \leq \sum_j |W_j| P_{jh}^k \quad \forall h \text{ and } k \quad (5)$$

Let us notate  $h_{\max}$  the index of the highest module  $|W_h|$ ,  $h = 1, n$

$$|\lambda^k| |W_{h_{\max}}| \leq |W_{h_{\max}}| \sum_j P_{jh}^k \leq |W_{h_{\max}}| \sum_j \max_j P_{jh}^k \leq |W_{h_{\max}}| n$$

thus  $|\lambda^k| \leq n$  for any  $k$  which implies that  $|\lambda| \leq 1$ . The diagonal matrix  $\Lambda$  can then be ordered by decreasing value of the magnitude of the latent values  $1 < |\lambda_2| < \dots < |\lambda_n|$ .

Because of the stochasticity of the transition of a Markov chain, Eq. (2) can be written

$$P^k = U\Lambda^k V^\top = U_1 V_1^\top + \sum_{i=2,n} \lambda_i^k U_i V_i^\top \quad (6)$$

The transition probabilities  $P_{ij}^k$  reach a limit  $V_{1j}$  ( $j = 1, n$ ) in the sense that  $\lim_{k \rightarrow \infty} P_{ij}^k = V_{1j}, \forall i$ . In order to prove it, it is sufficient to right multiply by  $P_{jh}$ , to sum over  $j$  and to use Eq. (3)

$$\lim_{k \rightarrow \infty} \sum_{j=1}^n P_{ij}^k P_{jh} = \sum_{j=1}^n V_{1j} P_{jh} = V_{1h}$$

We can also see from Eq. (6) that the speed of the convergence depends essentially on the module of second highest latent value  $|\lambda_2|$ .

### 1.3.1. Two states: activity and inactivity by age

In the simple case of two states, active and inactive, the transition matrix  $P$  can be written

$$P = \begin{pmatrix} 1-c & c \\ a & 1-a \end{pmatrix} = U\Lambda V^\top \quad (7)$$

$$\Lambda = \begin{pmatrix} 1 & 0 \\ 0 & 1-c-a \end{pmatrix} \quad (8)$$

$$U = \frac{1}{\sqrt{c+a}} \begin{pmatrix} 1 & -c \\ 1 & a \end{pmatrix} \quad (9)$$

$$V^\top = \frac{1}{\sqrt{c+a}} \begin{pmatrix} a & c \\ -1 & 1 \end{pmatrix} = U^{-1} \quad (10)$$

$$(c+a)P = \begin{pmatrix} a & c \\ a & c \end{pmatrix} + (1-c-a) \begin{pmatrix} c & c \\ -a & a \end{pmatrix} \quad (11)$$

$$(c+a)P^k = (c+a)U\Lambda^k V^\top = \begin{pmatrix} a & c \\ a & c \end{pmatrix} + (1-c-a)^k \begin{pmatrix} c & c \\ -a & a \end{pmatrix} \quad (12)$$

We suppose that  $0 < a < 1$  and  $0 < c < 1$ . We will study the case with absorbing states in a next section.

Let  $A_x$  and  $I_x$  be the number of people respectively active and inactive at age  $x$ ,  $N$  be the total population of age  $x$  or  $x+1$ ,  $N = A_x + I_x$ ,  $y_x$  the prevalence of activity  $y_x = A_x/N$ . Then, the numbers of people active and inactive at age  $x+1$  and  $x+k$  are given by the formula

$$\begin{pmatrix} A_{x+1} & I_{x+1} \end{pmatrix} = \begin{pmatrix} A_x & I_x \end{pmatrix} P \quad (13)$$

$$\begin{pmatrix} A_{x+k} & I_{x+k} \end{pmatrix} = \begin{pmatrix} A_x & I_x \end{pmatrix} P^k \quad (14)$$

$$(15)$$

In terms of prevalence of activity the formula is

$$\begin{pmatrix} y_{x+1} & 1 - y_{x+1} \end{pmatrix} = \begin{pmatrix} y_x & 1 - y_x \end{pmatrix} P \quad (16)$$

$$\begin{pmatrix} y_{x+k} & 1 - y_{x+k} \end{pmatrix} = \begin{pmatrix} y_x & 1 - y_x \end{pmatrix} P^k \quad (17)$$

and we can write

$$\lim_{k \rightarrow \infty} P^k = \begin{pmatrix} \frac{a}{a+c} & \frac{c}{a+c} \\ \frac{a}{a+c} & \frac{c}{a+c} \end{pmatrix} \quad (18)$$

or

$$\begin{pmatrix} A_{x+k} & I_{x+k} \end{pmatrix} \simeq \begin{pmatrix} A_x & I_x \end{pmatrix} \begin{pmatrix} \frac{a}{a+c} & \frac{c}{a+c} \\ \frac{a}{a+c} & \frac{c}{a+c} \end{pmatrix} \quad (19)$$

$$\lim_{k \rightarrow \infty} A_{x+k} = \frac{a}{a+c} (A_x + I_x) = \frac{a}{a+c} N + \quad (20)$$

$$+ \frac{(1-c-a)^k}{a+c} (cA_x - aI_x) \quad (21)$$

$$y_{x+k} = \frac{a}{a+c} + \frac{(1-c-a)^k}{a+c} \frac{cA_x - aI_x}{N} \quad (22)$$

$$y_{x+k} = \frac{a}{a+c} + (1-c-a)^k \left( y_x - \frac{a}{a+c} \right) \quad (23)$$

$$(24)$$

When  $k$  is big and because  $|\lambda_2| = |1-c-a| < 1$ , the prevalence of activity converges to a limit independent of the initial prevalence

$$\lim_{k \rightarrow \infty} y_{x+k} = y_L = \frac{a}{a+c} \quad (25)$$

We can also remark that when the initial prevalence at age  $x$  is the stationary prevalence  $y_x = y_L$ , the prevalence at any age  $x+k$  is constant and also equal to the stationary prevalence.

When  $1-c-a = 0$  the limit  $y_L = a = 1-c$  is reached in one step. When  $1-c-a \leq 0$  the convergence is alternated. When  $a+c$  is close to the unreachable (by definition) maximum of 2, the convergence to  $\frac{1}{2}$  is slow and alternated.

Let us now review how we classically define a retroprojection that is unrelated to what we call a backward prevalence.

### 1.3.2. Two states, retroprojection

As none of the  $n$  eigen values of  $P$  is equal to zero, we can calculate the number of people active and inactive at age  $x$  by reversing the recurrence and using the inverse  $P^{-1}$  of the transition matrix  $P$

$$\begin{pmatrix} y_x & 1 - y_x \end{pmatrix} = P^{-1} \begin{pmatrix} y_{x+1} & 1 - y_{x+1} \end{pmatrix} \quad (26)$$

$$P^{-1} = \frac{1}{1-a-c} \begin{pmatrix} 1-a & -c \\ -a & 1-c \end{pmatrix} \quad (27)$$

$$P^{-1} = U \Lambda^{-1} U^{-1} \quad (28)$$

or, with numbers instead of prevalence and after  $k$  iterations

$$\begin{pmatrix} A_x & I_x \end{pmatrix} = P^{-1} \begin{pmatrix} A_{x+1} & I_{x+1} \end{pmatrix} \quad (29)$$

$$\begin{pmatrix} A_{x-k} & I_{x-k} \end{pmatrix} = (P^{-1})^k \begin{pmatrix} A_{x+1} & I_{x+1} \end{pmatrix} \quad (30)$$

$$(c+a)P^{-k} = (c+a)U\Lambda^{-k}V^\top = \begin{pmatrix} a & c \\ a & c \end{pmatrix} + (1-c-a)^{-k} \begin{pmatrix} c & c \\ -a & a \end{pmatrix} \quad (31)$$

It is now obvious that the process is divergent and that the new prevalence is very sensitive to the prevalence value at age  $x+1$

$$\begin{pmatrix} A_{x-k} & I_{x-k} \end{pmatrix} \simeq \frac{1}{(1-c-a)^k} \begin{pmatrix} c & c \\ -a & a \end{pmatrix} \begin{pmatrix} A_{x+1} & I_{x+1} \end{pmatrix} \quad (32)$$

$$y_{x-k} = \frac{a}{a+c} + (1-c-a)^{-k} \left( y_{x+1} - \frac{a}{a+c} \right) \quad (33)$$

$$(34)$$

### 1.3.3. Two states, backprobability

In order to reconstruct an initial table,  $M$ , with numbers, similar to Table 1, we consider that the population is of size  $N$  (we simplify the notation  $N_x$ ) for all ages  $x$ .

The population at age  $x$  is divided into two subpopulations, the active people  $yN$  and the inactive people  $(1-y)N$ . Let  $c$  and  $a$  be the constant parameters of the forward matrix of transition  $P_x$ , then we can write

$$M(c, a, y, N) = \begin{pmatrix} (1-c)yN & cyN \\ a(1-y)N & (1-a)(1-y)N \end{pmatrix} \quad (35)$$

The active population at age  $x+1$  and year  $t+1$  is then  $(1-c)yN + a(1-y)N$  out of which  $a(1-y)N$  people were inactive at age  $x$  and year  $t$ . The backward probability for a woman to be inactive at age  $x$  knowing that she will be active at age  $x+1$  is then

$$\gamma = \frac{a(1-y)}{(1-c)y + a(1-y)} \quad (36)$$

Also the backward probability for a woman to have been active at age  $x$  knowing that she is inactive at age  $x+1$  is

$$\alpha = \frac{cy}{cy + (1-a)(1-y)} \quad (37)$$

The prevalence of activity at age  $x+1$  is

$$y^* = (1-c)y + a(1-y) = a + y(1-c-a) \quad (38)$$

is constant but differs from the prevalence  $y$  unless  $y = \frac{a}{a+c}$ .



With the notations  $\alpha$ ,  $\gamma$  and  $y^*$ , we can write  $M$  as

$$M(\alpha, \gamma, y^*, N) = \begin{pmatrix} (1-\gamma)y^*N & \alpha(1-y^*)N \\ \gamma y^*N & (1-\alpha)(1-y^*)N \end{pmatrix} \quad (39)$$

so that the transposed matrix  $M^t(\alpha, \gamma, y^*, N)$  has an analogous structure to  $M(c, a, y, N)$ .

The backward matrix of probabilities is thus dependent of the initial prevalence  $y$  (at age  $x$ ) and can be written

$$B_{x+1}(y) = \begin{pmatrix} 1-\gamma(y) & \alpha(y) \\ \gamma(y) & 1-\alpha(y) \end{pmatrix} \quad (40)$$

Its transposed matrix

$$B_{x+1}^\top(y) = \begin{pmatrix} 1-\gamma(y) & \gamma(y) \\ \alpha(y) & 1-\alpha(y) \end{pmatrix} \quad (41)$$

is a stochastic matrix of transition from age  $x+1$  to  $x$  which corresponds to a Markov chain converging into the past.

We can also write

$$M = N \begin{pmatrix} 1-\gamma & \alpha \\ \gamma & 1-\alpha \end{pmatrix} \begin{pmatrix} y^* & 0 \\ 0 & 1-y^* \end{pmatrix} = N \begin{pmatrix} y & 0 \\ 1-y & 1-y \end{pmatrix} \begin{pmatrix} 1-c & c \\ a & 1-a \end{pmatrix} \quad (42)$$

We can use the notation  $w_x^i$  for the prevalence in state  $i$  at age  $x$ , and here with only two states, activity and non activity, we have  $w_x^1 = y_x$ ,  $w_x^2 = 1 - y_x$ . Also using the notation  $W_x^d$  for the diagonal matrix of prevalences in each state, we have

$$W_x^d = \begin{pmatrix} w_x^1 & 0 \\ 0 & w_x^2 \end{pmatrix} = \begin{pmatrix} y_x & 0 \\ 0 & 1-y_x \end{pmatrix}. \quad (43)$$

Using matrices,  $M$  can be written

$$M = NB_{x+1}W_{x+1}^d = NW_x^dP_x \quad (44)$$

and the duality can be expressed in this form hereafter

$$B_{x+1}W_{x+1}^d = W_x^dP_x. \quad (45)$$

The matrix transition of the backward process, or backward Markov chain, is

$$B_{x+1}^\top = (W_{x+1}^d)^{-1}P_x^\top W_x^d \quad (46)$$

or

$$\begin{pmatrix} 1-\gamma & \gamma \\ \alpha & 1-\alpha \end{pmatrix} = \begin{pmatrix} \frac{1}{y_{x+1}} & 0 \\ 0 & \frac{1}{1-y_{x+1}} \end{pmatrix} \begin{pmatrix} 1-c & a \\ c & 1-a \end{pmatrix} \begin{pmatrix} y_x & 0 \\ 0 & 1-y_x \end{pmatrix} \quad (47)$$

from which the values of  $\gamma$  (Eq. 36) and  $\alpha$  (Eq.37) can be easily verified.

Introducing the logit function defined by  $\text{logit } x = \log \frac{x}{1-x}$ , we can write:

$$\begin{aligned}\frac{1}{\gamma} - 1 &= \frac{1-c}{a} \frac{y}{1-y} \\ \frac{1}{\alpha} - 1 &= \frac{1-a}{c} \frac{1-y}{y} \text{ and by taking the log, we get} \\ \text{logit } \gamma &= -\log\left(\frac{1-c}{a}\right) - \text{logit } y \\ \text{logit } \alpha &= -\log\left(\frac{1-a}{c}\right) + \text{logit } y\end{aligned}$$

From last two equations we get

$$\text{logit } y = \frac{1}{2} \left( \text{logit } \alpha - \text{logit } \gamma + \log \frac{(1-a)a}{(1-c)c} \right)$$

$$\text{logit } \gamma + \text{logit } \alpha = \text{logit } c + \text{logit } a$$

Also, in last equation, the sum of the two logits is independent of the initial prevalence  $y$ .

The left eigen vectors of the backward transition matrix  $B_{x+1}^\top$  must satisfy  $B_{x+1}^\top U^* = U^* \Lambda^*$  and the right eigen vector  $V^{*\top} B_{x+1}^\top = \Lambda^* V^{*\top}$ . Their values can be deduced from former equations involving  $c$ ,  $a$  and  $y$  (Eq. 7 and following).

$$\Lambda^* = \begin{pmatrix} 1 & 0 \\ 0 & 1 - \gamma - \alpha \end{pmatrix} \quad (48)$$

$$U^* = \frac{1}{\sqrt{\gamma + \alpha}} \begin{pmatrix} 1 & -\gamma \\ 1 & \alpha \end{pmatrix} \quad (49)$$

$$V^{*\top} = \frac{1}{\sqrt{\gamma + \alpha}} \begin{pmatrix} \alpha & \gamma \\ -1 & 1 \end{pmatrix} = U^{*-1} \quad (50)$$

$$B^\top = U^* \Lambda^* V^\top \quad (51)$$

$$(\gamma + \alpha)(B^\top) = \begin{pmatrix} \alpha & \gamma \\ \alpha & \gamma \end{pmatrix} + (1 - \gamma - \alpha) \begin{pmatrix} \gamma & \gamma \\ -\alpha & \alpha \end{pmatrix} \quad (52)$$

$$(\gamma + \alpha)(B^\top)^k = (\gamma + \alpha)U\Lambda^k V^\top = \begin{pmatrix} \alpha & \gamma \\ \alpha & \gamma \end{pmatrix} + (1 - \gamma - \alpha)^k \begin{pmatrix} \gamma & \gamma \\ -\alpha & \alpha \end{pmatrix} \quad (53)$$

Using the decomposition of Eq. (7) we can directly write the duality equation (Eq. 46)

$$P^\top = U^{-1\top} \Lambda U^\top = V \Lambda V^{-1} \quad (54)$$

$$B_{x+1}^\top = (W_{x+1}^d)^{-1} V \Lambda V^{-1} W_x^d \quad (55)$$

$$B_{x+1}^\top = ((W_{x+1}^d)^{-1} V) \Lambda ((W_x^d)^{-1} V)^{-1} \quad (56)$$

but the latter decomposition doesn't correspond to the diagonalisation. Also we can understand that the eigen values of  $B_{x+1}^\top$  won't be very different from those of

$P_x$  because in the real life the prevalences at age  $x + 1$ ,  $W_{x+1}^d$ , will be very close to the prevalences at age  $x$ ,  $W_x^d$ . Using our example of two states, the second eigen value of matrix  $P_x$ ,  $1 - c - a$  is different from the second eigen value of  $B_{x+1}^\top$ ,  $1 - \gamma - \alpha$ , because  $c + a \neq \gamma + \alpha$  but according to Eq. (48) the difference is slim and the equality is exact via the logit transformation.

All the results of the previous subsection are still valid: the prevalence at younger ages converges to a limit  $y_L^*$  verifying:

$$y_L^* = \frac{\alpha}{\alpha + \gamma}. \quad (57)$$

**Property 1.** *If the initial prevalence of the forward process corresponds to the stationary state, it corresponds also to the stationary prevalence of the backward process.*

PROOF. Replacing  $y$  by  $y_L = \frac{a}{a+c}$  in Eq. (36) and Eq. (37) we find that  $\gamma(y_L) = c$  and  $\alpha(y_L) = a$  so that  $y_L^* = \frac{\alpha}{\alpha+\gamma} = \frac{a}{a+c} = y_L$ .

We have also an obvious property

**Property 2.** *If the convergence to the forward stationary prevalence is alternated, the convergence to the backward stationary prevalence is also alternated.*

PROOF. The convergence of the forward process changes to alternated as soon as the second eigen value  $1 - a - c$  of the forward transition matrix is negative. When  $1 - a + c = 0$  in Eq. (36) and Eq. (37) we found that  $\gamma = 1 - y$  and  $\alpha = y$  so that the second eigen value of the backward transition  $1 - \alpha - \gamma$  is also equal to zero when the convergence starts to be alternated too.

#### 1.4. Weak ergodicity

In the previous section, we discussed the simple case when  $a_x$  and  $c_x$  were independent of the age  $x$ , but we have seen in the first part of this chapter that the main sociological interest resides in the fact that the probability to enter or reenter the labor force as well as the probability to leave the labor force are varying by age as shown in Fig. 3.

Despite the ups and downs that can be observed on Fig: 3 (and which are due to the fact that the numbers of people observed by single age are too few), the variations of the parameters  $a_x$  and  $c_x$  between successive ages are significative but small so that the product of a high number of transition matrices will slightly vary with age, and not converge to a fixed matrix with equal rows.

But an important phenomena will remain. This phenomena is called the weak ergodicity and its property is that the product matrix  ${}_k P_x = P_x P_{x+1} \cdots P_{x+k-1}$  tends to a matrix with equal rows. Rows are changing with  $k$ , but tend to be equal. It has been demonstrated in many different ways (Lopez, 1961), (Cohen, 1979) but H. Lebras's proof is particularly simple (Le Bras, 1971).

In terms of prevalence the formula 17 is now

$$\begin{pmatrix} y_{x+1} & 1 - y_{x+1} \end{pmatrix} = \begin{pmatrix} y_x & 1 - y_x \end{pmatrix} P_x \quad (58)$$

$$\text{or } \begin{pmatrix} y_{x+k} & 1 - y_{x+k} \end{pmatrix} = \begin{pmatrix} y_x & 1 - y_x \end{pmatrix} P_x P_{x+1} \cdots P_{x+k-1}. \quad (59)$$

It is useful to calculate the forward prevalence of activity starting from a cohort of only active women  $\begin{pmatrix} 1 & 0 \end{pmatrix}$  at a quinquennial age (as in Fig. 4) or inactive women  $\begin{pmatrix} 0 & 1 \end{pmatrix}$  in order to visualize how the stationary prevalence is bound by these two extreme upper ( $y^U$ ) and lower curves ( $y^L$ ) up to convergence

$$\begin{pmatrix} y_{x+k}^U & 1 - y_{x+k}^U \end{pmatrix} = \begin{pmatrix} 1 & 0 \end{pmatrix} {}_k P_x \quad (60)$$

$$\begin{pmatrix} y_{x+k}^L & 1 - y_{x+k}^L \end{pmatrix} = \begin{pmatrix} 0 & 1 \end{pmatrix} {}_k P_x \quad (61)$$

We can see that once the convergence is reached for a product  $k$  matrices, the convergence holds for any further projection. At the limit, we can write

$${}_k P_x = P_x P_{x+1} \cdots P_{x+k-1} \simeq \begin{pmatrix} {}_k y_{x+k} & 1 - {}_k y_{x+k} \\ {}_k y_{x+k} & 1 - {}_k y_{x+k} \end{pmatrix} \text{ or} \quad (62)$$

$$\simeq \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} {}_k y_{x+k} \\ 1 - {}_k y_{x+k} \end{pmatrix} \quad (63)$$

where  ${}_k y_{x+k}$  is the stationary prevalence of activity at age  $x + k$  computed with the product of  $k$  matrices since age  $x$ .

It is probably easier to use the notation  $y_x^\infty$  for the stationary prevalence of activity at age  $x$  (and  $1 - y_x^\infty$  for the prevalence of inactivity)

$$\begin{pmatrix} y_x^\infty & 1 - y_x^\infty \\ y_x^\infty & 1 - y_x^\infty \end{pmatrix} = \lim_{k \rightarrow \infty} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} {}_k P_{x-k} = \lim_{k \rightarrow \infty} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} P_{x-k} P_{x-k+1} \cdots P_{x-1}. \quad (64)$$

And as

$$\begin{pmatrix} y_x^\infty & 1 - y_x^\infty \\ y_x^\infty & 1 - y_x^\infty \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} y_x^\infty \\ 1 - y_x^\infty \end{pmatrix},$$

we get

$$\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} y_x^\infty \\ 1 - y_x^\infty \end{pmatrix} = \lim_{k \rightarrow \infty} {}_k P_{x-k} = \lim_{k \rightarrow \infty} P_{x-k} P_{x-k+1} \cdots P_{x-1}. \quad (65)$$

If the convergence is not reached with enough precision after a product of  $k$  matrices, the rows are not identical and we will keep both upper and lower bounds

$$\begin{pmatrix} {}_k y_x^U & 1 - {}_k y_x^U \\ {}_k y_x^L & 1 - {}_k y_x^L \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} {}_k P_{x-k} = {}_k P_{x-k}. \quad (66)$$

If the stationary prevalence is known up to age  $x$ , the stationary prevalence at age  $x + 1$  is simply given by the recurrent Eq. 17

$$\begin{pmatrix} y_{x+1}^\infty & 1 - y_{x+1}^\infty \end{pmatrix} = \begin{pmatrix} y_x^\infty & 1 - y_x^\infty \end{pmatrix} P_x \quad (67)$$

$$\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} y_{x+1}^\infty \\ 1 - y_{x+1}^\infty \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} y_x^\infty \\ 1 - y_x^\infty \end{pmatrix} P_x. \quad (68)$$

We must also consider a difficulty for young people. Indeed, in order to achieve convergence at a young age, say age 15, we need to start calculating the matrix product from a much younger age, something like age 10 or 5. It means that if data on flows were only available from age 15 to 74, we need to extrapolate the function  $c_x$  and  $a_x$  of Fig. 3 from ages 5 or 10 to age 15 in order to get the convergent prevalence at age 15 (see Fig. 1). In this example, it was easy to extrapolate the probabilities to enter the labor force before age 15 as something close to zero and the probabilities to leave the labor force as something very high. But if these assumptions were wrong, the stationary forward prevalence could not be accurately estimated. Therefore the recurrent Eq. (67) is useful if and only if the stationary prevalence is known at the first age  $x_0$  where the parameters of the transition matrix  $P_{x_0}$  (here  $c_{x_0}$  and  $a_{x_0}$ ) start to be known.

The series of left eigen vectors of each age specific transition matrix play an important role in the calculation of the stable prevalence but no easy formula could be found yet. Therefore, the stable prevalence has been calculated with the use of a computer in order to get the product of matrices.

The same conclusion applies to the age specific backward prevalence which cannot be calculated by easy formulas involving the latent left vectors of the age specific matrices of backprobabilities.

In the case of the backprevalence of activity at old ages, and in order to get convergence, we supposed that after age 74 it was not possible to enter the labor force ( $a_x = 0$ ) and that the probability to leave the labor force ( $c_x$ ) was close to 1.  $\gamma_{x+1}$  and  $\alpha_{x+1}$  have been extrapolated in the same way on both extremities.

The backward equation of recurrence is simply

$$\begin{pmatrix} y_x^{*\infty} & 1 - y_x^{*\infty} \end{pmatrix} = \begin{pmatrix} y_{x+1}^{*\infty} & 1 - y_{x+1}^{*\infty} \end{pmatrix} B_{x+1}^\top \quad (69)$$

$$\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} y_x^{*\infty} & 0 \\ 0 & 1 - y_x^{*\infty} \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} y_{x+1}^{*\infty} & 0 \\ 0 & 1 - y_{x+1}^{*\infty} \end{pmatrix} B_{x+1}^\top \quad (70)$$

and when the convergence is not reachable, we have to deal with the upper and lower back prevalences, starting at age  $x + k$  to get prevalences at age  $x$  implying  $k$  backward transitions matrices

$$\begin{pmatrix} ky_x^{*U} & 1 - ky_x^{*U} \\ ky_x^{*L} & 1 - ky_x^{*L} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} {}_{-k}B_{x+k}^\top = {}_{-k}B_{x+k}^\top = B_{x+k}^\top B_{x+k-1}^\top \cdots B_{x+1}^\top. \quad (71)$$

We can write an equation similar to Eq. (45)

$$M = N \cdot B_{x+1}^O W_{x+1,t+1}^{dO} = N \cdot W_x^{dO} P_x^O \quad (72)$$

$$B_{x+1}^O W_{x+1,t+1}^{dO} = W_{x,t}^{dO} P_x^O \quad (73)$$

where  $W_{x+1,t+1}^{dO} = \text{diag}(w_{x+1,t+1}^{1O}, w_{x+1,t+1}^{2O}, \dots, w_{x+1,t+1}^{iO}, \dots)$  is the diagonal matrix of the prevalence in state  $i$  at age  $x+1$  and date  $t+1$ . We use the superscript “ $O$ ” when the data are observed from the survey.

### 1.5. Backward prevalence of a specific cohort

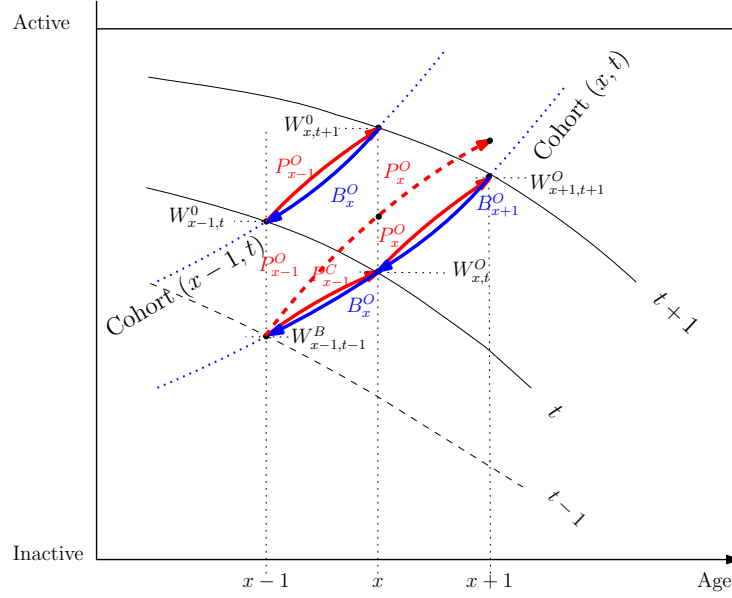


Figure 7: Zoom on trajectories of two cohorts  $(x, t)$  and  $(x-1, t)$  and their observed prevalence in activity at first survey in  $t$  and second survey in  $t+1$ .

We represented in Fig. 7 a detail of some trajectories:

- the observed prevalences (in activity and in inactivity) of the cohort  $(x, t)$  observed at the first round at  $t$  and age  $x$ ,  $W_{x,t}^O$ , and at the second round at  $t+1$  and age  $x+1$ ,  $W_{x+1,t+1}^O$ ;
- the same observed trajectory of the younger cohort  $(x-1, t)$ ,  $W_{x-1,t}^O$  and  $W_{x,t+1}^O$ .

From the observed prevalences of the  $(x, t)$  cohort at  $t$  and  $t+1$  we deduced the forward matrix of transition  $P_x^O$  and the backward matrix  $B_{x+1}^O$ , and from the observed prevalences of the  $(x-1, t)$  cohort at  $t$  and  $t+1$  we deduced the forward matrix of transition  $P_{x-1}^O$  and the backward matrix  $B_x^O$ .

Eq. (73) can be written under the elementary form

$$b_{x+1}^{ijO} w_{x+1}^{jO} = w_x^{iO} p_x^{ijO} \quad \forall i, j, \quad (74)$$

and thus, the Markov forward projection of the observed prevalence at  $(x, t)$  by  $P_x^O$

$$W_{x+1,t+1}^\top = \begin{pmatrix} w_{x+1,t+1}^1 & w_{x+1,t+1}^2 \end{pmatrix}^\top = \begin{pmatrix} w_{x,t}^{1O} & w_{x,t}^{2O} \end{pmatrix}^\top P_x^O \quad (75)$$

which also can be written under the elementary form

$$w_{x+1}^j = \sum_i w_{x,t}^{iO} p_x^{ijO} \quad \forall j, \quad (76)$$

is equal the observed prevalence at age  $x + 1$ . The proof is obtained by simple replacement

$$w_{x+1}^j = \sum_i b_{x+1}^{ijO} w_{x+1}^{jO} = w_{x+1}^{jO} \quad \forall j. \quad (77)$$

and because the transposed matrix  $B_{x+1}^{O\top}$  is stochastic and  $\sum_i b_{x+1}^{ijO} = 1, \forall j$ .

Under the assumption that the age specific observed backward prevalences,  $B_x^O$ , are constant over time, we can deduce the backward prevalence of the cohort  $(x, t)$  at age  $x - 1$  by applying the Markov backward projection  $B_x^O$  to the observed prevalence at age  $x$  ( $W_{x,t}^O$ ).

But the trick here is that the process is no more reversible. In fact, if the pair  $(B_x^O, P_{x-1}^O)$  applied to  $W_{x,t+1}^O$ , moved to  $W_{x-1,t}^O$  and  $W_{x,t+1}^O$  again, the same pair  $(B_x^O, P_{x-1}^O)$  applied to  $W_{x,t}^O$  moves to the so-called backward prevalence at age  $x - 1$ ,  $W_{x-1,t-1}^B$ , but then to a prevalence at age  $x$  which differs from  $W_{x,t}^O$ . The reason comes from the fact that

$$B_x^O W_{x,t}^{dO} \neq W_{x-1,t-1}^{dB} P_{x-1}^O \quad (78)$$

even if

$$B_x^O W_{x,t+1}^{dO} = W_{x-1,t}^{dO} P_{x-1}^O. \quad (79)$$

It reminds us that by changing the prevalence level at age  $x$  and keeping the backward matrix unchanged, the forward matrix at age  $x$  can't remain constant and *has to be changed* to a matrix  $P_{x-1}^C$ , satisfying equation

$$B_x^O W_{x,t}^{dO} = W_{x-1,t-1}^{dB} P_{x-1}^C \quad (80)$$

in order to move back to the same cohort.

In Fig. 7, we applied this constant forward matrix  $P_{x-1}^O$  but also applied  $P_x^O$  to the result. It can be seen that this trajectory (red dashed) is different from the trajectory of the cohort  $(x, t)$ .

Having this remark in mind, we can now calculate the backward prevalences of the same cohort  $(x, t)$  at any age  $x - k$  by chaining the constant backward matrices  $B_x^{OT}$  into a matrix  $(_{-k}B_x^O)^\top$  and we get  $(w_{x-k,t-k}^{1B} \quad w_{x-k,t-k}^{2B})^\top$

$$(W_{x-k,t-k}^B)^\top = (W_{x+1,t+1}^O)^\top (_{-(k+1)}B_{x+1}^O)^\top \quad (81)$$

$$= (W_{x+1,t+1}^O)^\top B_{x+1}^{O\top} B_x^{O\top} B_{x-1}^{O\top} \cdots B_{x-k+1}^{O\top} \quad (82)$$

$$= (W_{x,t}^O)^\top B_x^{O\top} B_{x-1}^{O\top} \cdots B_{x-k+1}^{O\top} \quad (83)$$

$$= (W_{x,t}^O)^\top (_{-k}B_x^O)^\top \quad (84)$$

with

$$(_{-k}B_x^B)^\top = B_x^{O\top} B_{x-1}^{O\top} \cdots B_{x-k+1}^{O\top}. \quad (85)$$

In the case of  $J$  states, the above equations are still valid with

$$W_{x-k,t-k}^B = \begin{pmatrix} w_{x-k}^{11} & w_{x-k}^{12} & \dots & w_{x-k}^{1J} \end{pmatrix} \quad (86)$$

$$W_{x,t}^O = \begin{pmatrix} w_x^{11O} & w_x^{12O} & \dots & w_x^{1JO} \end{pmatrix} \quad (87)$$

$$(88)$$

### 1.6. Forward prevalence of a specific cohort

Similarly, assuming that the observed forward transition matrices at any age  $x$ ,  $P_x^O$  are constant over time, we can deduce the forward prevalences of the cohort  $(x, t)$  in any state  $i$  and age  $x + k$

$$W_{x+k,t+k}^F = W_{x,t}^O \cdot {}_kP_x^F = W_{x+1,t+1}^O \cdot {}_{k-1}P_{x+1}^F \quad (89)$$

with

$${}_kP_x^F = P_x^O P_{x+1}^O \dots P_{x+k-1}^O \quad (90)$$

## 2. Backward probabilities with transient states and an absorbing state

As in the previous section, we will take an easy example in order to present the methodology before entering the theoretical part.

Using data from the Italian SILC survey (Giudici et al., 2017) which is a follow up study with two interviews on the health status (active or inactive), we get the following table

		Act.(1)	In.(2)	Dead
Active(1)	9983	7998	1932	53
Inactive	6607	1828	4525	254
Total	16590	9826	6457	307

This table can also be represented on the Lexis diagrams of Fig. 8. The classical approach using forward probabilities is on the right. From a cohort of 9983 people active at age  $x$ , only 7998 were still active (state 1) at age  $x + 1$ , 1932 were inactive at the date of the second interview (state 2) and 4 died (state 3) before the second interview. The same process is applied to the 6457 inactive people at the first interview. If we consider a Markov chain model with 3 states, active, inactive and dead, the various probabilities are represented on the figure.

On the left side of the same Fig.8 we defined the backward probabilities. Starting from the 9826 people who survived and were active at the second interview, we know that 7998 were active (1) and 1828 were inactive (2) at the first interview. Thus,  $b_{21} = \frac{1828}{9826}$  is the probability to be inactive (1) a year before, knowing the current status (2) at age  $x$  and  $b_{11} = \frac{7998}{9826}$  etc.

In the forward approach, we computed  $p_{13} = \frac{53}{9983}$  as the probability for an active person to die between the two interviews. In the backward approach, even



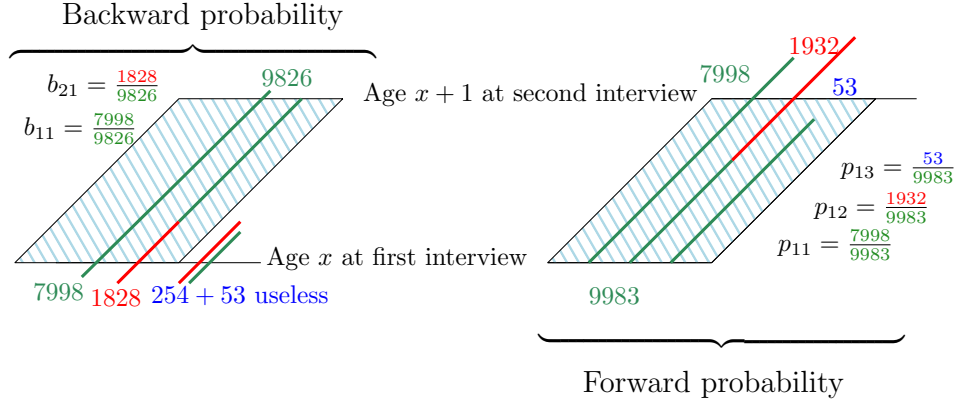


Figure 8: Lexis diagram of forward and backward approaches with DIFFERENTIAL MORTALITY. Green for active (state 1), red for inactive (2), blue for death (3).

if we could define  $b_{13} = \frac{53}{307}$  as the probability, knowing that we will die between the two interviews, to be active at the first interview, we are only interested in past history of a cohort of survivors,

It could be argued that backward prevalences involve the past history of a very selected population who by chance survived to old ages but it is not completely true because this backward cohort is starting with few people at old ages but its size increases at younger ages and is very numerous at very young ages.

Also the same argument could apply to forward prevalences of a classical life table where probability of death at old ages is estimated from a highly selected population and is not reflecting the mortality of the total population if that population was reaching these old ages.

Let us now move to the theoretical part using the case with only three alive states (1, 2, 3) but with an absorbing state (4). But it is not difficult to generalize to an arbitrary number of non absorbing states. This section involves only matrices and is presented here because our IMaCh software which is able to treat an arbitrary number of alive states, uses such kind of matrix products. We will add the superscript “e” to distinguish the full extended matrices from the matrices reduced to the transient states.

Let us first consider the  $M$  matrix which describes how  $N^{\cdot\cdot}$  people in various alive states  $i$  at age  $x$  are found at age  $x + 1$  in state  $j$  at the next wave.

Sometimes we will remove the  $x$  index for better readability.  $N_{x+1}^4 = N^{14} + N^{24} + N^{34}$  people died between the two waves.

$$\begin{matrix} N_x^1. \\ N_x^2. \\ N_x^3. \\ N^{\cdot\cdot} \end{matrix} \begin{pmatrix} N^{11} & N^{12} & N^{13} & N^{14} \\ N^{21} & N^{22} & N^{23} & N^{24} \\ N^{31} & N^{32} & N^{33} & N^{34} \end{pmatrix} \begin{matrix} N_{x+1}^1 \\ N_{x+1}^2 \\ N_{x+1}^3 \\ N_{x+1}^4 \end{matrix}$$

Using an arbitrary last row, we can rewrite  $M$  as a squared matrix

$$\begin{aligned}
 M &= \begin{pmatrix} N^{11} & N^{12} & N^{13} & N^{14} \\ N^{21} & N^{22} & N^{23} & N^{24} \\ N^{31} & N^{32} & N^{33} & N^{34} \\ 0 & 0 & 0 & N^{\cdot\cdot} \end{pmatrix} \\
 &= \begin{pmatrix} N_x^{1\cdot} & 0 & 0 & 0 \\ 0 & N_x^{2\cdot} & 0 & 0 \\ 0 & 0 & N_x^{3\cdot} & 0 \\ 0 & 0 & 0 & N^{\cdot\cdot} \end{pmatrix} \begin{pmatrix} p^{11} & p^{12} & p^{13} & p^{14} \\ p^{21} & p^{22} & p^{23} & p^{24} \\ p^{31} & p^{32} & p^{33} & p^{34} \\ 0 & 0 & 0 & 1 \end{pmatrix} \\
 M &= \begin{pmatrix} b^{11} & b^{12} & b^{13} & b^{24} \\ b^{21} & b^{22} & b^{23} & b^{24} \\ b^{31} & b^{32} & b^{33} & b^{34} \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} N_{x+1}^{1\cdot} & 0 & 0 & 0 \\ 0 & N_{x+1}^{2\cdot} & 0 & 0 \\ 0 & 0 & N_{x+1}^{3\cdot} & 0 \\ 0 & 0 & 0 & N^{\cdot\cdot} \end{pmatrix} \quad (91)
 \end{aligned}$$

$$= B_{x+1}^e \text{diag}(N_{x+1}^{1\cdot}, N_{x+1}^{2\cdot}, N_{x+1}^{3\cdot}, N^{\cdot\cdot}) \quad (92)$$

$$M = N^{\cdot\cdot} \begin{pmatrix} b^{11} & b^{12} & b^{13} & b^{24} \\ b^{21} & b^{22} & b^{23} & b^{24} \\ b^{31} & b^{32} & b^{33} & b^{34} \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} w_{x+1}^{1eO} & 0 & 0 & 0 \\ 0 & w_{x+1}^{2eO} & 0 & 0 \\ 0 & 0 & w_{x+1}^{3eO} & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad (93)$$

$$= N^{\cdot\cdot} B_{x+1}^e \text{diag}(w_{x+1}^{1eO}, w_{x+1}^{2eO}, w_{x+1}^{3eO}, 1) \quad (94)$$

$$M = N^{\cdot\cdot} \begin{pmatrix} w_x^{1O} & 0 & 0 & 0 \\ 0 & w_x^{2O} & 0 & 0 \\ 0 & 0 & w_x^{3O} & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} p^{11} & p^{12} & p^{13} & p^{14} \\ p^{21} & p^{22} & p^{23} & p^{24} \\ p^{31} & p^{32} & p^{33} & p^{34} \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad (95)$$

$$= N^{\cdot\cdot} \cdot \text{diag}(w_x^{1O}, w_x^{2O}, w_x^{3O}, 1) \cdot P_x \quad (96)$$

where  $w_x^{iO} = \frac{N_x^{i\cdot}}{N^{\cdot\cdot}}$  is the observed prevalence in state  $i$  at age  $x$  and  $p^{ij} = \frac{N_x^{ij}}{N_x^{i\cdot}}$  is the forward probability to be in state  $j$  ( $j = 1, \dots, 4$ ) at  $x + 1$ , being in state  $i$ , ( $i = 1, \dots, 3$ ), at age  $x$ .  $w_{x+1}^{jeO} = \frac{N_{x+1}^{j\cdot}}{N^{\cdot\cdot}}$  is the extended observed prevalence at age  $x + 1$  in state  $j$  including death, while  $w_{x+1}^{jO} = \frac{N_{x+1}^{j\cdot}}{N_{x+1}^{\cdot\cdot} - N_{x+1}^{4\cdot}}$  is the observed prevalence among the survivors at age  $x + 1$  in state  $j$ .

The link between the extended matrices  $B_{x+1}^e$  and  $P_x^e$  with the extended prevalences at  $x + 1$  and the observed prevalences at  $x$  is given by the following equation, similar to Eq. (73)

$$B_{x+1}^e \text{diag}(w_{x+1}^{1eO}, w_{x+1}^{2eO}, \dots, 1) = \text{diag}(w_x^{1O}, w_x^{2O}, \dots, 1) \cdot P_x^e. \quad (97)$$

### 2.1. Chaining forward for a specific cohort

As the set  $(P_x^{eO}, \forall x)$  is a set of stochastic matrices, we can suppose that they are constant over time and chain them in the same manner that with did in section 1.6 and using Eq. (90) in order get the forward extended prevalences at age  $x + 1$  and year  $t + 1$  from the observed prevalences at age  $x$  and year  $t$

$$\begin{pmatrix} w_{x+1,t+1}^{1eF} & w_{x+1,t+1}^{2eF} & w_{x+1,t+1}^{3eF} & 0 \end{pmatrix} = \begin{pmatrix} w_{x,t}^{1O} & w_{x,t}^{2O} & w_{x,t}^{3O} & 0 \end{pmatrix} \cdot P_x^e \quad (98)$$

$$(W_{x+1,t+1}^{eF})^\top = (W_{x,t}^O)^\top \cdot P_x^{eO} \quad (99)$$

And we can chain forward up to age  $x + k$

$$(W_{x+2,t+2}^{eF})^\top = (W_{x,t}^O)^\top P_x^{eO} P_{x+1}^{eO} = (W_{x,t}^O)^\top \cdot {}_2P_x^{eO} \quad (100)$$

$$(W_{x+k,t+k}^{eF})^\top = (W_{x,t}^O)^\top P_x^{eO} P_{x+1}^{eO} \cdots P_{x+k-1}^{eO} \quad (101)$$

$$(W_{x+k,t+k}^{eF})^\top = (W_{x,t}^O)^\top \cdot {}_kP_x^{eO} \quad (102)$$

In the case of three alive transient states and death, it can be written

$$\begin{pmatrix} w_{x+k}^{1eF} & w_{x+k}^{2eF} & w_{x+k}^{3eF} & w_{x+k}^{4eF} \end{pmatrix} = \begin{pmatrix} w_{x,t}^{1O} & w_{x,t}^{2O} & w_{x,t}^{3O} & 0 \end{pmatrix} \begin{pmatrix} {}_kP_x^{11} & {}_kP_x^{12} & {}_kP_x^{13} & {}_kP_x^{14} \\ {}_kP_x^{21} & {}_kP_x^{22} & {}_kP_x^{23} & {}_kP_x^{24} \\ {}_kP_x^{31} & {}_kP_x^{32} & {}_kP_x^{33} & {}_kP_x^{34} \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad (103)$$

We can also rewrite the full extended matrix  ${}_kP_x^{eO}$  in order to display the submatrix of the transient state only

$${}_kP_x^{eO} = \begin{pmatrix} {}_kP_x^{O} & e - {}_kP_x^{O} \\ 0 & 1 \end{pmatrix} \quad e = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \quad (104)$$

$${}_kP_x^{O} = \begin{pmatrix} {}_kP_x^{11} & {}_kP_x^{12} & {}_kP_x^{13} \\ {}_kP_x^{21} & {}_kP_x^{22} & {}_kP_x^{23} \\ {}_kP_x^{31} & {}_kP_x^{32} & {}_kP_x^{33} \end{pmatrix} \quad (105)$$

Even if the matrix  ${}_kP_x^{O}$  is not stochastic, the property of ergodicity applies to this submatrix and when  $k$  increases, its rows tend to be identical. The matrix changes with the horizon  $k$  but is unique and depends only on the flows  $(p_y^{ij} \forall y \geq x)$  and no more on the prevalences at age  $x$ .

The reader may refer to a complete proof, similar to Lebras's proof, in the appendix of our article (Lièvre et al., 2003).

We can remark that the vector of the observed prevalence at age  $x$  is a sum of three matrices

$$\begin{pmatrix} w_{x,t}^{1O} & w_{x,t}^{2O} & w_{x,t}^{3O} & 0 \end{pmatrix} = w_{x,t}^{1O} \begin{pmatrix} 1 & 0 & 0 & 0 \end{pmatrix} + w_{x,t}^{2O} \begin{pmatrix} 0 & 1 & 0 & 0 \end{pmatrix} + w_{x,t}^{3O} \begin{pmatrix} 0 & 0 & 1 & 0 \end{pmatrix} \quad (106)$$

Therefore we can focus on the behavior of each of them

$$\begin{pmatrix} w_{x+k}^{11eF} & w_{x+k}^{12eF} & w_{x+k}^{13eF} & w_{x+k}^{14eF} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \end{pmatrix} {}_kP_x^{eO} \quad (107)$$

$$= \begin{pmatrix} 1 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} {}_kP_x^{11} & {}_kP_x^{12} & {}_kP_x^{13} & {}_kP_x^{14} \\ {}_kP_x^{21} & {}_kP_x^{22} & {}_kP_x^{23} & {}_kP_x^{24} \\ {}_kP_x^{31} & {}_kP_x^{32} & {}_kP_x^{33} & {}_kP_x^{34} \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad (108)$$

$$= \begin{pmatrix} {}_kP_x^{11} & {}_kP_x^{12} & {}_kP_x^{13} & {}_kP_x^{14} \end{pmatrix} \quad (109)$$

$$\begin{pmatrix} w_{x+k}^{21eF} & w_{x+k}^{22eF} & w_{x+k}^{23eF} & w_{x+k}^{24eF} \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 & 0 \end{pmatrix} {}_kP_x^{eO} = \begin{pmatrix} {}_kP_x^{21} & {}_kP_x^{22} & {}_kP_x^{23} & {}_kP_x^{24} \end{pmatrix} \quad (110)$$

$$\begin{pmatrix} w_{x+k}^{31eF} & w_{x+k}^{32eF} & w_{x+k}^{33eF} & w_{x+k}^{34eF} \end{pmatrix} = \begin{pmatrix} 0 & 0 & 1 & 0 \end{pmatrix} {}_kP_x^{eO} = \begin{pmatrix} {}_kP_x^{31} & {}_kP_x^{32} & {}_kP_x^{33} & {}_kP_x^{34} \end{pmatrix} \quad (111)$$

or of its summary

$$\begin{pmatrix} w_{x+k}^{11eF} & w_{x+k}^{12eF} & w_{x+k}^{13eF} & w_{x+k}^{14eF} \\ w_{x+k}^{21eF} & w_{x+k}^{22eF} & w_{x+k}^{23eF} & w_{x+k}^{24eF} \\ w_{x+k}^{31eF} & w_{x+k}^{32eF} & w_{x+k}^{33eF} & w_{x+k}^{34eF} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} {}_kP_x^{11} & {}_kP_x^{12} & {}_kP_x^{13} & {}_kP_x^{14} \\ {}_kP_x^{21} & {}_kP_x^{22} & {}_kP_x^{23} & {}_kP_x^{24} \\ {}_kP_x^{31} & {}_kP_x^{32} & {}_kP_x^{33} & {}_kP_x^{34} \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad (112)$$

and we draw on Fig. 9 the 9 elements of matrix  ${}_kP_x$ , starting at age  $x = 55$ , up to an increasing age  $x = 55 + k$ . In addition, we visualized the survival curves in each of the 3 states,  $s^i = 1 - p^{i4} = p^{i1} + p^{i2} + p^{i3}$ .

We can see that the mortality of a person living in an institution is much more important than the mortality of a person living in coresidence and is similar to that of a person living alone.

But because of the high mortality at these old ages, all the 9 elements are collapsing to zero and the ergodicity is not obvious.

Also for each of three starting values  $i$ , we got the result that the forward extended prevalences  $w_{x+k}^{ijeF}$  in state  $j$  at age  $x + k$  are simply the element  $(i, j)$  of the extended matrix  ${}_kP_x^{eO}$ ,  ${}_kP_x^{ij}$ .

The forward prevalences,  $w_{x+k}^{ijF}$ , are simply obtained by dividing by the probabilities of survival in  $k$  years being initially in state  $i$  at age  $x$ . This probability being  $1 - {}_kP_x^{i4}$ , the forward prevalences are

$$w_{x+k}^{ijF} = \frac{{}_kP_x^{ij}}{{}_kP_x^{i1} + {}_kP_x^{i2} + {}_kP_x^{i3}} \quad (113)$$

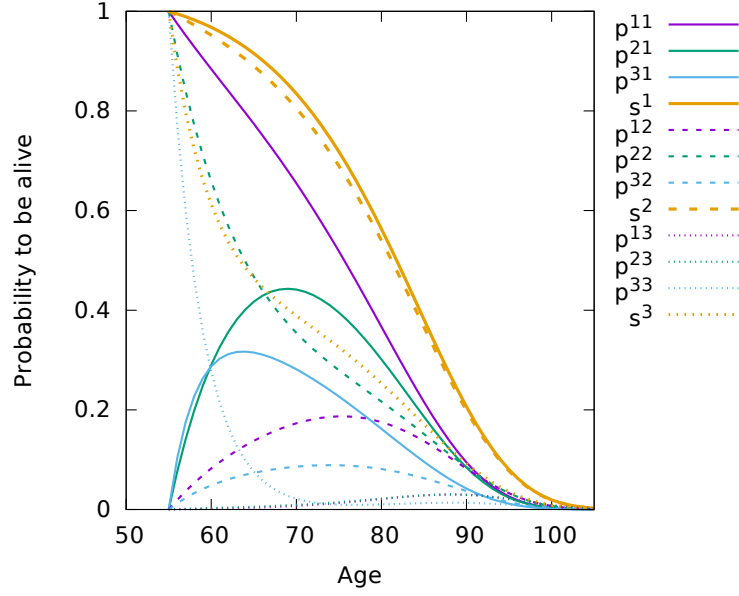


Figure 9: Survival functions for a person living initially at age 55 either in coresidence ( $i = 1$ ), alone (2) or in an institution (3). Probability  ${}_{x-55}p_{55}^{ij}$  to be at age  $x$  alive in state  $j = 1, 2$  or 3 and in any state ( $s_i$ ), being in state  $i$  at age 55. HRS (1998-2014).

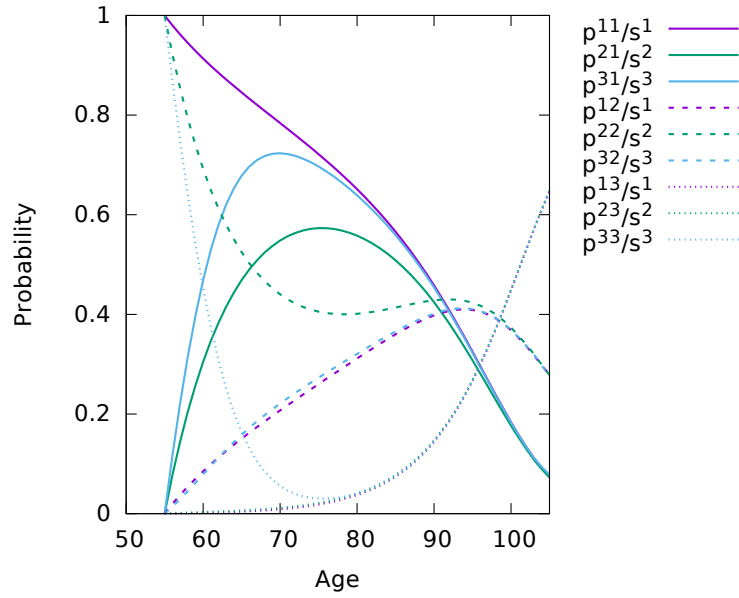


Figure 10: Identical to Fig. 9 but divided by probability to survive  ${}_{x-55}p_{55}^{ij}/{}_{x-55}s_{55}^i$ . Convergence to the forward prevalences.

In Fig. 10, we divided each row of the matrix by the corresponding probability to survive and now, the nine elements are clearly converging to three different curves which are the unique forward prevalences in each of the three states, coresidence, alone and in institutions.

The extended prevalences of a cohort is then a weighted mean of the three state specific prevalences and the forward prevalences are the results of their division by the overall probability to survive.

From Eq. 103, we can write  $w_{x+k}^{4e}$  as the overall probability to die within  $k$  years when the population at age  $x$  is distributed in state  $i$  according to  $w_x^{iO}$

$$w_{x+k}^{4e} = w_x^{1O} {}_k p_x^{14} + w_x^{2O} {}_k p_x^{24} + w_x^{3O} {}_k p_x^{34} \quad (114)$$

$$= 1 - (w_{x+k}^{1eF} + w_{x+k}^{2eF} + w_{x+k}^{3eF}) \quad (115)$$

and its opposite, the probability to survive is a weighted mean of the various probabilities of survive when originally in a state

$${}_k s_x^{wO} = 1 - w_{x+k}^{4eF} \quad (116)$$

$$= \sum_i w_x^{iO} {}_k s_x^i \quad (117)$$

$$= w_x^{1O} ({}_k p_x^{11} + {}_k p_x^{12} + {}_k p_x^{13}) + w_x^{2O} ({}_k p_x^{21} + {}_k p_x^{22} + {}_k p_x^{23}) + \quad (118)$$

$$w_x^{3O} ({}_k p_x^{31} + {}_k p_x^{32} + {}_k p_x^{33}). \quad (119)$$

The prevalences at age  $x + k$  are simply the proportions of people in each state among the survivors at age  $x + k$

$$w_{x+k}^{jF} = \frac{w_{x+k}^{jeF}}{{}_k s_x^{wO}} = w_x^{1O} \cdot \frac{{}_k p_x^{1j}}{{}_k s_x^{wO}} + w_x^{2O} \cdot \frac{{}_k p_x^{2j}}{{}_k s_x^{wO}} + w_x^{3O} \cdot \frac{{}_k p_x^{3j}}{{}_k s_x^{wO}}. \quad (120)$$

Using matrices, we can write

$$(W_{x+k,t+k}^F)^\top = (W_{x+k,t+k}^{eF})^\top \frac{1}{{}_k s_x^{wO}} = (W_{x,t}^O)^\top \cdot \frac{1}{{}_k s_x^{wO}} \cdot {}_k P_x^{eO} = (W_{x,t}^O)^\top \overline{{}_k P_x^{eO}} \quad (121)$$

with

$$\overline{{}_k P_x^{eO}} = \frac{1}{{}_k s_x^{wO}} \cdot {}_k P_x^{eO}, \quad (122)$$

We also have the following recurrences

$$\overline{{}_k P_x^{eO}} = \overline{{}_{k-1} P_x^{eO}} \cdot \overline{{}_k P_{x+k-1}^{eO}} \quad (123)$$

$$(W_{x+k,t+k}^F)^\top = (W_{x+k-1,t+k-1}^F)^\top \overline{{}_k P_{x+k-1}^{eO}} \quad (124)$$

with

$$\overline{{}_k P_{x+k-1}^{eO}} = \frac{{}_k P_{x+k-1}^{eO}}{{}_{k-1} \sigma_x^{wO}} \quad (125)$$

$$\overline{{}_k P_x^{eO}} = \frac{{}_k P_x^{eO}}{\sigma_x^{wO}} \quad (126)$$

and

$${}_k s_x^i = {}_{k-1} s_x^i \cdot s_{x+k-1}^i \quad (127)$$

$${}_{k-1} \sigma_x^{wO} = \frac{{}_k s_x^{wO}}{{}_{k-1} s_x^{wO}} = \sum_i w_x^{iO} \cdot s_{x+k-1}^i \frac{{}_{k-1} s_x^i}{\sum_j w_x^{jO} \cdot {}_{k-1} s_x^j} \cdot \quad (128)$$

With the Eq. (124), we are back to an inhomogenous Markov chain without absorbing state but this Eq. (124) differs from Eq. (59) because the coefficient  ${}_{k-1} \sigma_x^{wO} = \frac{{}_k s_x^{wO}}{{}_{k-1} s_x^{wO}}$  depends on the level of the prevalences at age  $x$  and not only on the elements of the last matrix  $P_{x+k-1}^{eO}$ . We could have suspected that the forward prevalences were not unique but we saw that they are not dependent on both observed and state-specific initial prevalences.

If we had used forward Markov matrices that would have been resized into a stochastic matrix at each age

$$\widetilde{P_{x+k-1}^{eO}} = \frac{P_{x+k-1}^{eO}}{s_{x+k}^{wO}} \quad (129)$$

$$\widetilde{{}_k P_x^{eO}} = \widetilde{P_x^{eO}} \cdot \widetilde{P_{x+1}^{eO}} \cdots \widetilde{P_{x+k-1}^{eO}} \quad (130)$$

we would haven't found a unique forward prevalence unless the mortality from any state was identical at any age. And we just saw in the above example that it is usually false. For disability versus disability-free, the differential is about 4 at age 70 and only 1.5 at age 95.

In order to calculate the prevalence at age  $x$ , we could start from any observed prevalence  $w_{x-k}^{iO}$  in state  $i$  at age  $x - k$  or from a set of prevalences  $(W_{x-k}^O)^\top$ .

Hence transposing equation Eq. (120) and using the notation  ${}_{-k} w_x^{jF}$ , the observed prevalence in state  $j$  at age  $x$  is given by the formula

$${}_{-k} w_x^{jF} = \frac{{}_{-k} w_x^{jeF}}{{}_k s_{x-k}^w} = w_{x-k}^{1O} \cdot \frac{{}_k P_{x-k}^{1j}}{{}_k s_{x-k}^w} + w_{x-k}^{2O} \cdot \frac{{}_k P_{x-k}^{2j}}{{}_k s_{x-k}^w} + w_{x-k}^{3O} \cdot \frac{{}_k P_{x-k}^{3j}}{{}_k s_{x-k}^w} \cdot \quad (131)$$

Using matrices, it follows

$$({}_{-k} W_{x,t}^F)^\top = ({}_{-k} W_{x,t}^{eF})^\top \frac{1}{{}_k s_{x-k}^{wO}} = (W_{x-k,t-k}^O)^\top \cdot \frac{1}{{}_k s_{x-k}^{wO}} \cdot P_{x-k}^{eO} = (W_{x-k,t-k}^O)^\top \overline{{}_k P_{x-k}^{eO}} \quad (132)$$

with

$$\overline{{}_k P_{x-k}^{eO}} = \frac{1}{{}_k s_{x-k}^{wO}} \cdot {}_k P_{x-k}^{eO} \cdot \quad (133)$$

Thus, the unique forward prevalence at age  $x$  is the limit of  ${}_{-k} w_x^{jF}$  when  $k$  increases to infinity. Using Eq. (112) we can simply remark the unique forward

prevalences at age  $x$  in any state  $j$  are given by any column of the matrix  $\overline{{}_k P_{x-k}^{eO}}$  when  $k$  increases to infinity.

In practice, we compute the matrix  ${}_k P_{x-k}^{eO}$  which is the product of the  $k$  matrices  ${}_k P_{x-k}^{eO} {}_k P_{x-k+1}^{eO} \dots {}_k P_{x-2}^{eO} {}_k P_{x-1}^{eO}$ . Remembering Eq. (112), each row  $i$  corresponds to the extended forward prevalences by states. Thus a simple division of each element  $(i, j)$  of the row  $i$  by the sum of the first elements (which is probability  ${}_k s_{x-k}^i$  to survive until age  $x$  when in state  $i$  at age  $x - k$ ,  $1 - {}_k P_{x-k}^{i4}$ ), will output the forward prevalences in each state. Any row should output the same results.

In fact, the software checks that for each column, the relative difference between the maximum and the minimum is less than a given tolerance. When this tolerance is reached, the forward prevalence is obtained.

And thus we can write the forward prevalences using an expression which is only dependent of the  ${}_k P_{x-k}^{ij}$  and not of any observed prevalence

$${}_k S_{x-k}^d = \text{diag} \left( {}_k P_{x-k}^{11} + {}_k P_{x-k}^{12} + {}_k P_{x-k}^{13}, \right. \quad (134)$$

$$\left. {}_k P_{x-k}^{21} + {}_k P_{x-k}^{22} + {}_k P_{x-k}^{23}, \right. \quad (135)$$

$$\left. {}_k P_{x-k}^{31} + {}_k P_{x-k}^{32} + {}_k P_{x-k}^{33} \right) \quad (136)$$

$$= \begin{pmatrix} {}_k s_{x-k}^1 & 0 & 0 \\ 0 & {}_k s_{x-k}^2 & 0 \\ 0 & 0 & {}_k s_{x-k}^3 \end{pmatrix} \quad (137)$$

$$W_x^{F\infty\top} = \begin{pmatrix} w_x^{1F} & w_x^{2F} & w_x^{3F} \\ w_x^{1F} & w_x^{2F} & w_x^{3F} \\ w_x^{1F} & w_x^{2F} & w_x^{3F} \end{pmatrix} = \lim_{k \rightarrow \infty} ({}_k S_{x-k}^d)^{-1} \cdot {}_k P_{x-k} \quad (138)$$

$$W_x^{F\infty\top} = \lim_{k \rightarrow \infty} ({}_k S_{x-k}^d)^{-1} {}_k P_{x-k} = \lim_{k \rightarrow \infty} ({}_k S_{x-k}^d)^{-1} P_{x-k} P_{x-k+1} \dots P_{x-1} \quad (139)$$

Also, applying Eq. (124) with  $k = 0$  we get the recurrent equation

$$(W_{x,t}^F)^\top = (W_{x-1,t-1}^F)^\top \overline{{}_k P_{x-1}^{eO}} = (W_{x-1,t-1}^F)^\top \frac{{}_k P_{x-1}^{eO}}{\sigma_{x-1}^{wO}} \quad (140)$$

## 2.2. Chaining backward for a specific cohort

Let us now review similar results for the backward prevalence.

The backward probability  $b_{x+1}^{ij}$  at age  $x + 1$  is defined as below

$$b_{x+1}^{ij} = \frac{N^{ij}}{N_{x+1}^j} \quad i = 1, 3 \quad j = 1, 4 \quad (141)$$

with

$$N_{x+1}^j = \sum_{i=1}^3 N^{ij} = N^\cdot \sum_{i=1}^3 w_x^i p^{ij} = w_{x+1}^{je} N^\cdot, \quad j = 1, 4 \quad (142)$$

$$\text{or } w_{x+1}^{je} = \sum_{i=1}^3 w_x^i p^{ij} = w_x^1 p^{1j} + w_x^2 p^{2j} + w_x^3 p^{3j}. \quad (143)$$



According to Eq; (97) and using a matrix notation, we can write

$$B_{x+1}^e = \text{diag}(w_x^1, w_x^2, w_x^3, 1) P_x^e (\text{diag}(w_{x+1}^{1e}, w_{x+1}^{2e}, w_{x+1}^{3e}, 1))^{-1} \quad (144)$$

or

$$B_{x+1}^e = \begin{pmatrix} w_x^1 & 0 & 0 & 0 \\ 0 & w_x^2 & 0 & 0 \\ 0 & 0 & w_x^3 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} p^{11} & p^{12} & p^{13} & p^{14} \\ p^{21} & p^{22} & p^{23} & p^{24} \\ p^{31} & p^{32} & p^{33} & p^{34} \\ 0 & 0 & 0 & 1 \end{pmatrix}. \quad (145)$$

$$\begin{pmatrix} \frac{1}{w_x^1 p^{11} + w_x^2 p^{21} + w_x^3 p^{31}} & 0 & 0 & 0 \\ 0 & \frac{1}{w_x^1 p^{12} + w_x^2 p^{22} + w_x^3 p^{32}} & 0 & 0 \\ 0 & 0 & \frac{1}{w_x^1 p^{13} + w_x^2 p^{23} + w_x^3 p^{33}} & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad (146)$$

$$B_{x+1}^e = \begin{pmatrix} \frac{w_x^1 p^{11}}{w_x^1 p^{11} + w_x^2 p^{21} + w_x^3 p^{31}} & \frac{w_x^1 p^{12}}{w_x^1 p^{12} + w_x^2 p^{22} + w_x^3 p^{32}} & \frac{w_x^1 p^{13}}{w_x^1 p^{13} + w_x^2 p^{23} + w_x^3 p^{33}} & \frac{w_x^1 p^{14}}{w_x^1 p^{14} + w_x^2 p^{24} + w_x^3 p^{34}} \\ \frac{w_x^2 p^{21}}{w_x^1 p^{11} + w_x^2 p^{21} + w_x^3 p^{31}} & \frac{w_x^2 p^{22}}{w_x^1 p^{12} + w_x^2 p^{22} + w_x^3 p^{32}} & \frac{w_x^2 p^{23}}{w_x^1 p^{13} + w_x^2 p^{23} + w_x^3 p^{33}} & \frac{w_x^2 p^{24}}{w_x^1 p^{14} + w_x^2 p^{24} + w_x^3 p^{34}} \\ \frac{w_x^3 p^{31}}{w_x^1 p^{11} + w_x^2 p^{21} + w_x^3 p^{31}} & \frac{w_x^3 p^{32}}{w_x^1 p^{12} + w_x^2 p^{22} + w_x^3 p^{32}} & \frac{w_x^3 p^{33}}{w_x^1 p^{13} + w_x^2 p^{23} + w_x^3 p^{33}} & \frac{w_x^3 p^{34}}{w_x^1 p^{14} + w_x^2 p^{24} + w_x^3 p^{34}} \\ 0 & 0 & 0 & 1 \end{pmatrix}. \quad (147)$$

The arbitrary last row is used so that the product of matrices used in our IMaCh software, concerns only full square matrices of the same size. But we are only interested in the 3 by 3 sub-matrix of the alive states

$$B_{x+1} = \begin{pmatrix} w_x^1 & 0 & 0 \\ 0 & w_x^2 & 0 \\ 0 & 0 & w_x^3 \end{pmatrix} \begin{pmatrix} p^{11} & p^{12} & p^{13} \\ p^{21} & p^{22} & p^{23} \\ p^{31} & p^{32} & p^{33} \end{pmatrix}. \quad (148)$$

$$\begin{pmatrix} \frac{1}{w_x^1 p^{11} + w_x^2 p^{21} + w_x^3 p^{31}} & 0 & 0 \\ 0 & \frac{1}{w_x^1 p^{12} + w_x^2 p^{22} + w_x^3 p^{32}} & 0 \\ 0 & 0 & \frac{1}{w_x^1 p^{13} + w_x^2 p^{23} + w_x^3 p^{33}} \end{pmatrix} \quad (149)$$

$$B_{x+1} = \begin{pmatrix} \frac{w_x^1 p^{11}}{w_x^1 p^{11} + w_x^2 p^{21} + w_x^3 p^{31}} & \frac{w_x^1 p^{12}}{w_x^1 p^{12} + w_x^2 p^{22} + w_x^3 p^{32}} & \frac{w_x^1 p^{13}}{w_x^1 p^{13} + w_x^2 p^{23} + w_x^3 p^{33}} \\ \frac{w_x^2 p^{21}}{w_x^1 p^{11} + w_x^2 p^{21} + w_x^3 p^{31}} & \frac{w_x^2 p^{22}}{w_x^1 p^{12} + w_x^2 p^{22} + w_x^3 p^{32}} & \frac{w_x^2 p^{23}}{w_x^1 p^{13} + w_x^2 p^{23} + w_x^3 p^{33}} \\ \frac{w_x^3 p^{31}}{w_x^1 p^{11} + w_x^2 p^{21} + w_x^3 p^{31}} & \frac{w_x^3 p^{32}}{w_x^1 p^{12} + w_x^2 p^{22} + w_x^3 p^{32}} & \frac{w_x^3 p^{33}}{w_x^1 p^{13} + w_x^2 p^{23} + w_x^3 p^{33}} \end{pmatrix}. \quad (150)$$

which is equivalent to Eq. (45)

$$B_{x+1} = \text{diag}(w_x^i) P_x \text{diag} \left( \frac{1}{\sum_i w_x^i p_x^{ij}} \right) \quad (151)$$

The transposed matrix  $(B_{x+1})^\top$  (in fact, only its 2 by 2 submatrix of alive states)

$$(B_{x+1})^\top = \text{diag}\left(\frac{1}{\sum_i w_x^i p_x^{ij}}\right) P_x^\top \text{diag}(w_x^i) \quad (152)$$

is stochastic and corresponds to the transition matrix of the backward Markov chain within the alive states.

We can then right multiply the matrices to get the limit of the product of backward matrices

$$(-_k B_{x+k})^\top = (B_{x+k})^\top (B_{x+k-1})^\top \cdots (B_{x+2})^\top (B_{x+1})^\top. \quad (153)$$

In order to highlight the backward prevalences  $w_x^{iB}$  and forward prevalences  $w_x^{iF}$  at exact age  $x$  as a limit when  $k \rightarrow \infty$ , and using the notations generalizing to  $J$  states

$$(w_x^B)^\top = \begin{pmatrix} w_x^{1B} & w_x^{2B} & \cdots & w_x^{JB} \end{pmatrix} \quad (154)$$

$$(W_x^B)^\top = \begin{pmatrix} w_x^{1B} & w_x^{2B} & \cdots & w_x^{JB} \\ \vdots & \vdots & & \vdots \\ w_x^{1B} & w_x^{2B} & \cdots & w_x^{JB} \end{pmatrix} \quad (155)$$

we can write

$$(W_x^{B\infty})^\top = \lim_{k \rightarrow \infty} (-_k B_{x+k})^\top = \lim_{k \rightarrow \infty} (B_{x+k})^\top (B_{x+k-1})^\top \cdots (B_{x+2})^\top (B_{x+1})^\top \quad (156)$$

so that with Eq. (140) we get the generalized recurrent equations

$$(W_x^B)^\top = (W_{x+1}^B)^\top (B_{x+1})^\top \quad (157)$$

$$(W_x^F)^\top = (W_{x-1}^F)^\top \overline{P_{x-1}^{eO}} = (W_{x-1}^F)^\top \frac{P_{x-1}^{eO}}{\sigma_{x-1}^{wO}}. \quad (158)$$

### 2.2.1. Stationary population, forward and backward prevalences

Let us study again the simple example of the labor force and the activity ratio of women in France. We suppose that the mortality is neglectable or, and this is similar, that there is no differential of mortality between active and inactive women.

Let us consider a cohort, named  $\mathcal{F}$ , of  $N$  young women of age 14 born at year  $t$  and a set of matrices  $M$  similar to Table 1 for each age between 14 to 74 years. In order to simplify, we suppose that the size of the cohort at each age is constant and equal to  $N^\cdot = N$ .

We can set the age-specific transition matrices  $P_x$  using the series of  $c_x$ ,  $a_x$  observed in France in 1977-78 from  $x = 14$  to 74 years, calculate the number of active and inactive women as well as the stationary forward prevalence  $y_x^\infty$  of this cohort  $\mathcal{F}$  at each age  $x$ .

Let us suppose now that the population is stationary in the sense that each year, the number of young women reaching the age of 14 during any year  $t$  is constant

and equal to  $N$ . We also suppose that the age-specific transition matrices  $P_x$  are constant over time. It is easy to understand that the cross-sectional prevalence of this stationary population at any time  $t$  is constant and equal to the forward prevalence at each age. Also, this prevalence,  $y_x^\infty$  satisfies the recurrent Eq. 67.

On the other way, we can consider a cohort, named  $\mathcal{B}$ , of  $N$  old women of age 75 and compute the backward transition matrices  $B_{x+1}$  (series of  $\gamma_x, \alpha_x$  for  $x = 15$  to 75 years) at each age to get the number of active and inactive women at each younger age as well as the stationary backward prevalence  $y_x^{*\infty}$  which must satisfy Eq. 69.

Are there conditions that must satisfy  $c_x, a_x, \gamma_x$  and  $\alpha_x$  in order to produce equal forward and backward stable prevalences?

From Eq. 67 and 69 or, better, from generalized Eq. 157 and Eq. 158 at age  $x$ , we get

$$\begin{cases} (W_{x+1}^F)^\top &= (W_x^F)^\top \overline{P_x} \\ (W_x^B)^\top &= (W_{x+1}^B)^\top (B_{x+1})^\top. \end{cases} \quad (159)$$

If we assume the equality at age  $x + 1$ ,  $(W_{x+1}^F)^\top = (W_{x+1}^B)^\top$  we get

$$(W_x^B)^\top = (W_x^F)^\top \overline{P_x} (B_{x+1})^\top \quad (160)$$

so that the equality at any age  $x$  imposes the necessary following condition

$$(W_x^B)^\top = (W_x^F)^\top \overline{P_x} (B_{x+1})^\top \quad (161)$$

which doesn't seem to make any sense.

Let us now investigate a different approach which will result in the following theorem.

**Theorem 1.** *If the observed prevalence in an cohort of survivors is equal, for any state and age, to the stationary backward prevalence, the stationary forward prevalence is also equal to both of them.*

**PROOF OF THEOREM 1.** We can consider a cohort of old people,  $\mathcal{B}$ , whose transitions between states and two adjacent ages  $x + 1$  and  $x$  are known. Thus, we can calculate the age specific observed prevalences in any state as well as the stationary backward prevalences,  $(W_x^B)^\top$  constructed for example, by recurrence using the backward generalized Eq. 157 and starting from a very old age. Let us suppose that the observed backward prevalence is identical at any age and state  $i$  to the corresponding stationary backward prevalence. We could say that this old cohort  $\mathcal{B}$  is “backward stationary”.

Let us now calculate the forward transitions,  $\overline{P_x(\mathcal{B})}$ , as well as the forward prevalences  $(W_x^F(\mathcal{B}))^\top$  of this backward stationary cohort  $\mathcal{B}$  in any state  $i$  using the generalized forward recurrent Eq. 158 while starting yet from a very young age onward.

We will use a mathematical induction to prove that this forward prevalence  $(W_x^F(\mathcal{B}))^\top$  is equal to the backward prevalence  $(W_x^B)^\top$ .

Making the hypothesis that the theorem is true up to age  $x$ , we will prove that it is also true at age  $x + 1$ . More precisely let us suppose that this forward prevalence

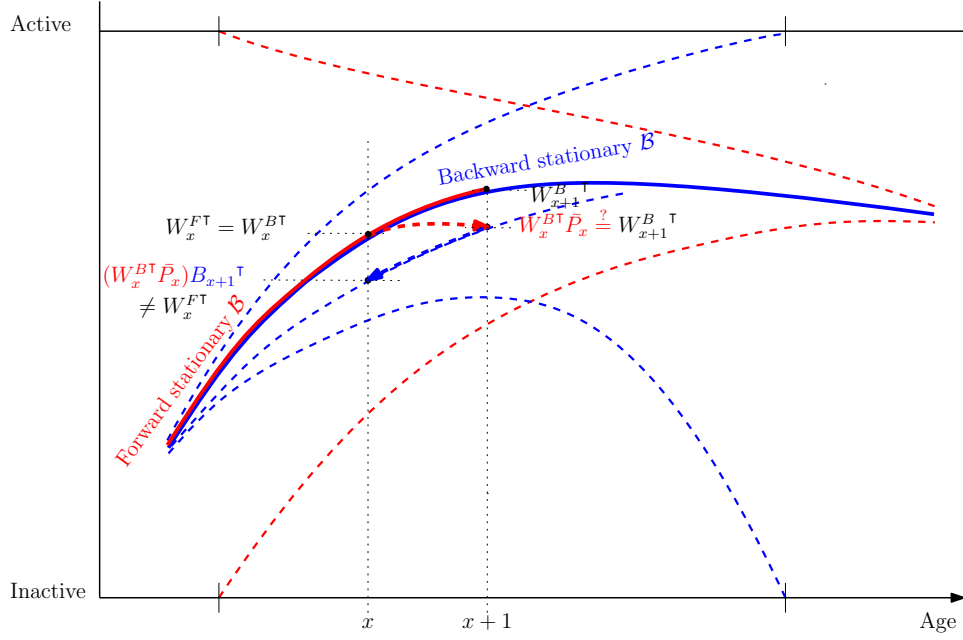


Figure 11: Help for the proof of theorem 1. Considering a backward stationary cohort  $\mathcal{B}$ , its associated forward prevalence should be identical to the backward prevalence. Blue is for backward projections, red for forward projections. Dotted lines are examples of trajectories either from activity or inactivity.

is equal to the backward prevalence starting from this very young age up to age  $x$  (Fig. 11)

$$(W_y^F)^\top = (W_y^B)^\top \quad \forall y \leq x. \quad (162)$$

We can then use Eq. 158 and above Eq. 162 at age  $x-1$  to calculate the forward prevalence at age  $x+1$ ,  $(W_{x+1}^F)^\top = (W_x^F)^\top \bar{P}_x = (W_x^B)^\top \bar{P}_x$ .

And finally, we can certify using a proof by contradiction that this calculated forward prevalence  $(W_x^B)^\top \bar{P}_x$  is identical to the backward stationary prevalence  $(W_{x+1}^B)^\top$ : In fact, if  $(W_x^B)^\top \bar{P}_x$  was not equal to  $(W_{x+1}^B)^\top$ , the resulting prevalence calculated from the backward recurrent Eq. 157 at age  $x$ ,  $((W_x^B)^\top \bar{P}_x)(B_{x+1})^\top$ , would differ from the prevalence  $(W_x^B)^\top$  and that is in contradiction with our former hypothesis of Eq. 162.

Also, because each of the two directions used, first forward and then backward, are directions which are convergins, we won't have precision issues.

This completes the mathematical induction and achieves the proof of the theorem.

A similar theorem applies to a cohort whose observed prevalences are equal to the stationary forward prevalences: the backward prevalences are identical to the observed and forward prevalences.

This theorem is important in practice because it means that if the cross-sectional prevalences observed during a first pass are identical to the prevalences deduced from the matrix of the forward transitions measured at a second pass, the prevalences deduced from the backward transitions must also be identical.

It can be said differently: if the forward prevalences are close to the cross-sectional prevalences, the backward prevalences are also close.

To be convinced and highlight the previous theorem, we drew in Fig. 13 and 12 the forward (violet) and backward (green) probabilities from the chain survey on the French female labor force (in bold lines) and their associated stationary series (in fine lines). A series and its associated will produce identical forward and backward prevalences.

As the situation in France was very far from stationnarity, the forward and backward probabilities were very different. But their associated probabilities, either forward or backward are closer from each other.

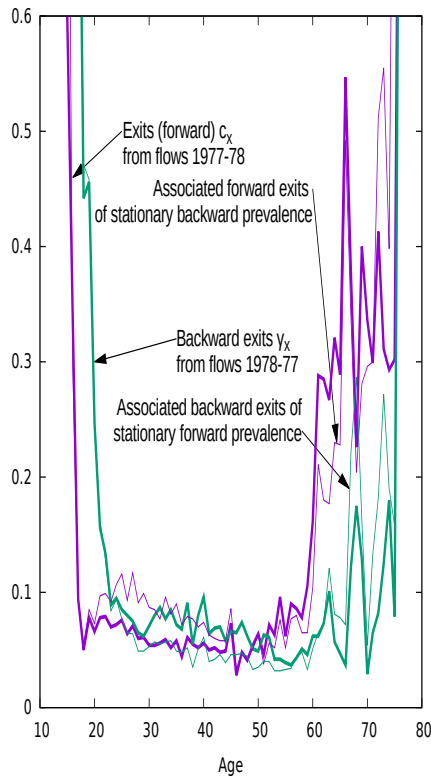


Figure 12: Exits: initial forward  $c_x$  and backward  $\gamma_x$  probabilities as well as their associated  $c_x(\mathcal{B})$  and  $\gamma_x(\mathcal{F})$ .

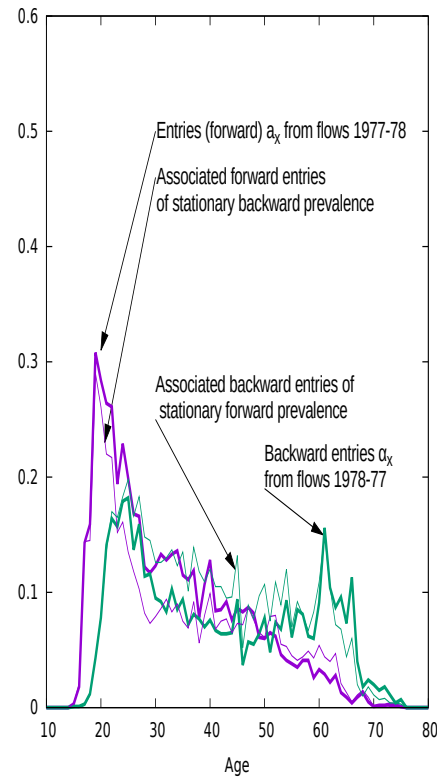


Figure 13: Entries: initial forward  $a_x$  and backward  $\alpha_x$  probabilities as well as their associated  $a_x(\mathcal{B})$  and  $\alpha_x(\mathcal{F})$ .

In Demography, it is a good practice to use a fictitious cohort in order to emphasize the changes in mortality or fertility rates observed during a short period of time. But in this chapter, we are introducing the concept of a fictitious cohort of old people whose characteristics were measured when they were younger.

The backward prevalences will reflect the behavior of the population in the past and the forward prevalences the behavior of the population in the future. The cross-sectional prevalences reflect the current situation mixing old and young people.

These three concepts are represented on the Lexis diagram of Fig. 14.

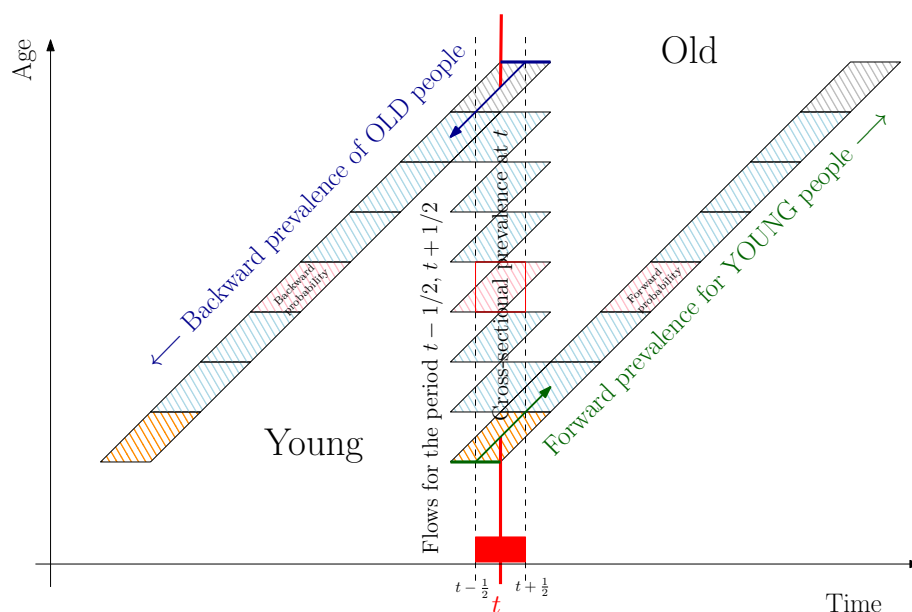


Figure 14: Age-specific forward probabilities and age-specific backward probabilities linked for young, respectively old cohorts.

On this figure, we can see a “census” vertical line at time  $t$  and a period of time between  $t - \frac{1}{2}$  and  $t + \frac{1}{2}$ . Usually, age specific rates or transitions are estimated in a square between two exact ages, but there values are similar to those estimated in a “perspective” diamond centered at the same mean date and mean age. These forward probabilities or rates belonging to a diamond can be moved to the right to correspond to what a “young” cohort could expect in its future if the rates remained constant. If we use mortality rates, we can compute the survival function that this young cohort could expect as well as its life expectancy.

The introduction of backward probabilities or backward rates estimated in the same diamonds is new. These diamonds and their corresponding backward values can be moved to the left in order to belong to a same cohort of “old” people. Unfortunately the backward probability of mortality is useless but for multistate models estimated from cross-longitudinal surveys, the concept of backward probability is meaningful, and, for example, backward prevalence in a health state can be estimated in an old cohort at each age.

The calculation of backward prevalences is therefore an important contribution to highlight the behavioral changes of the human populations: if a significant change in behavior is initiated, the backward prevalence will generally be in the opposite direction of the forward prevalence relative to cross-sectional prevalence.

Before the presentation of some results concerning these newly backward prevalences, it seems important to explain, using the analysis of mortality, how big the difference between a cross-sectional index and a forward or period index can be.

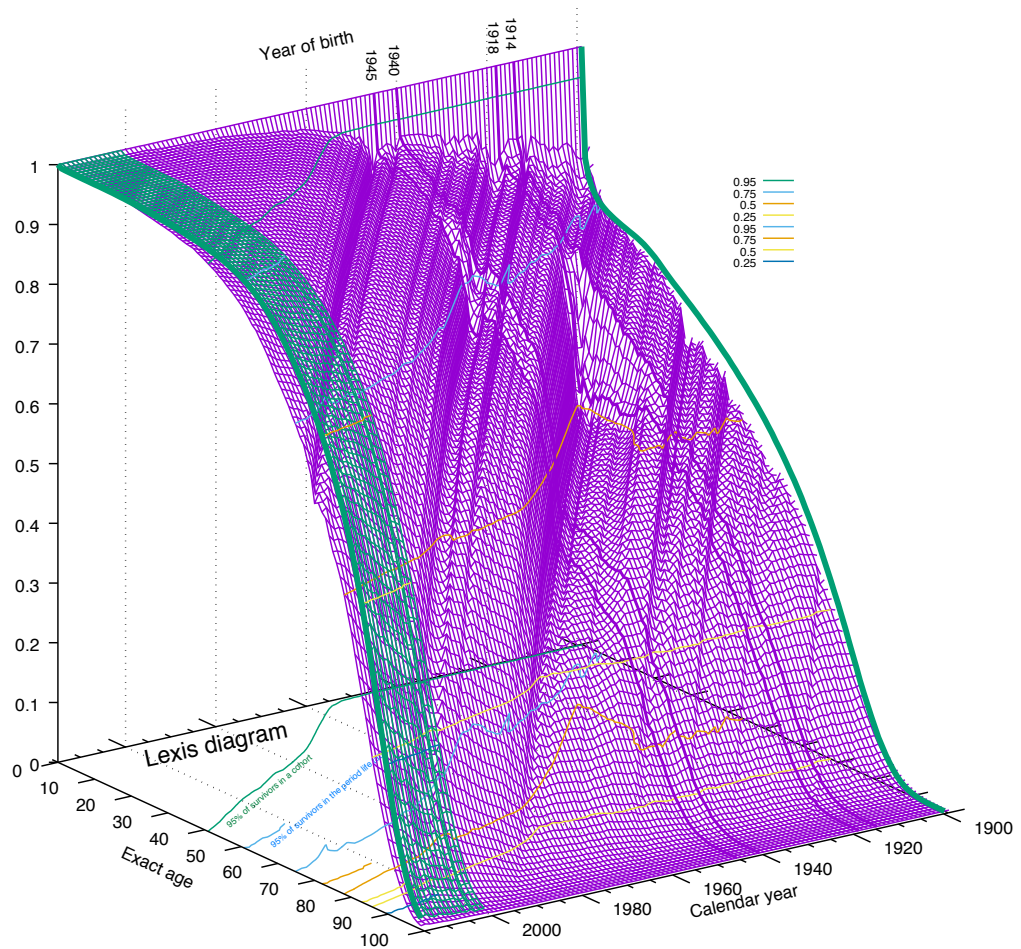


Figure 15: Proportion of male survivors from birth in France observed between 1899 and 2015, by exact age and calendar year. Depletions due to world wars I and II. Comparison with the annual period life tables since year 2000 and in 1899 (green surfaces). Contour plots at 95% (5% loss) of cohorts surface (green line) and period surface (blue line) are also reported on the Lexis diagram of the basis.

Let us imagine a stationary population, i.e a closed population with a constant number of births per unit of time and constant age specific mortality rates. In each cohort, the survival function is constant and its area under the curve or life expectancy is constant too.

Let us now assume that the number of births is still constant and equal to 1 per unit of time but that the mortality corresponded to the historical observed mortality by age since more than one and half century (see Fig. 15).

By analogy with the cross-sectional prevalence of disability or of any characteristic which can be measured during a census or a cross-sectional survey, we will name this “observed” proportion of survivors since birth, the “cross-sectional prevalence of life”.

Fig. 16 shows the cross-sectional prevalences of survival by gender in France

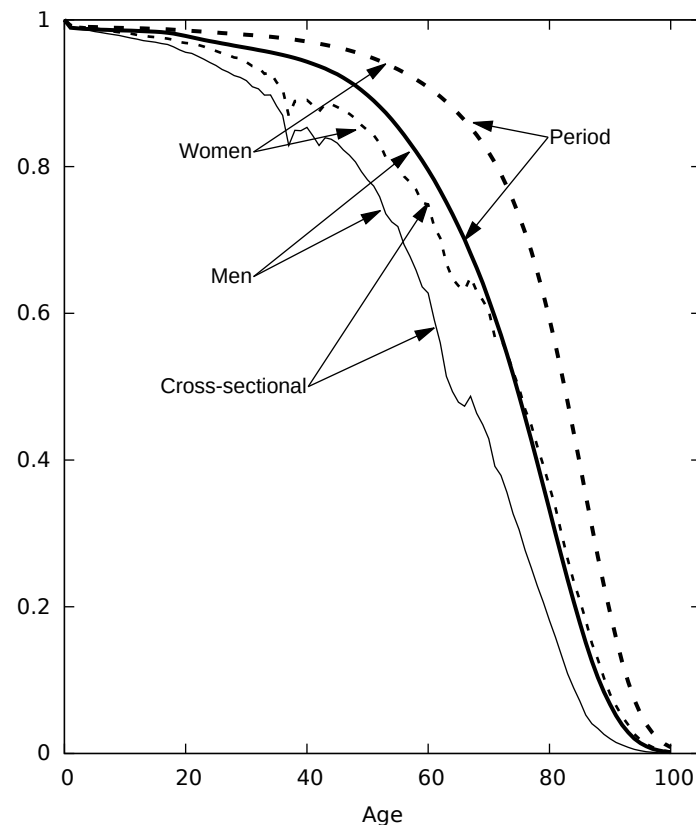


Figure 16: Cross-sectional prevalences of survival by gender in France in 1982. Comparison with the period life tables in 1982.

on mid 1982 when this concept has been first published (Brouard, 1986). They can be compared with the corresponding period life tables deduced from the mortality rates observed in France during 1982.

If the area under the period life table is the classical life expectancy at birth



$e_0$ , the area under the cross-sectional life prevalence, which I sometimes named  $d_0$  (for duration), is a “crosse-sectional” index (Brouard, 1986), now so-called the C.A.L index, for cross-sectional average length of life (Guillot, 2003), (Goldstein and Wachter, 2006), (Luy, 2006), (Canudas-Romo and Guillot, 2015).

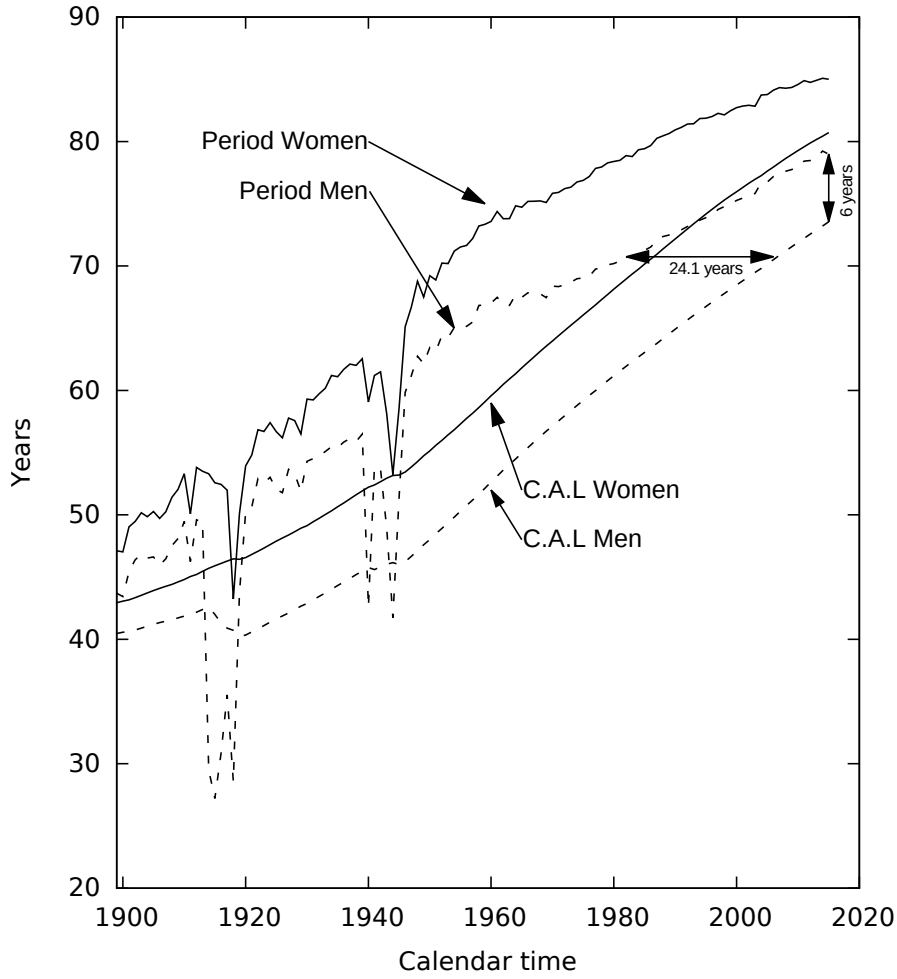


Figure 17: Cross-sectional duration of survival or C.A.L ( $d_0$ ) and Life expectancy ( $e_0$ ) in France by gender from 1899 to 2015.

In France, the difference between both indices which is visualized in Fig. 17 is still high (about 6 years for men and 5 years for women in 2015), reflecting the fact that the current age structure of the French (male) population is still much younger than what the low current mortality levels will impose. These days, the difference is getting smaller because the larger gap was due to the historic decline in infant mortality and young adults, which could be seen until 1950 (Fig. 15). After 1950, infant mortality was already low, so even though it continued to decline in relative variation, absolute variations were small, minimizing the tempo effect. We can see

in this figure that according to the 2015 period life table, 95% of French men could reach the age of 50, but in reality within any male cohort ever born in France, 5% already died before the age of 40 (95% of the most recent cohort was only 40 years old).

The tempo effect was aggravated by world wars, particularly because of the tragic mortality caused by the First World War of 1914-1918 among men (Fig. 15).

In 1982, assuming that the mortality rates observed in 1982 would remain constant in the future, it would have been possible to predict the cross-sectional prevalence of survival, each year to come, by projecting survival of each cohort under this hypothesis of mortality constancy.

Such predictions for 1992, 2002, 2012, 2022 and 2032 are represented on Fig. 18. The convergence to be the period prevalence, or period life table of 1982, is almost achieved by 2022. The last comparison, which is of great interest, con-

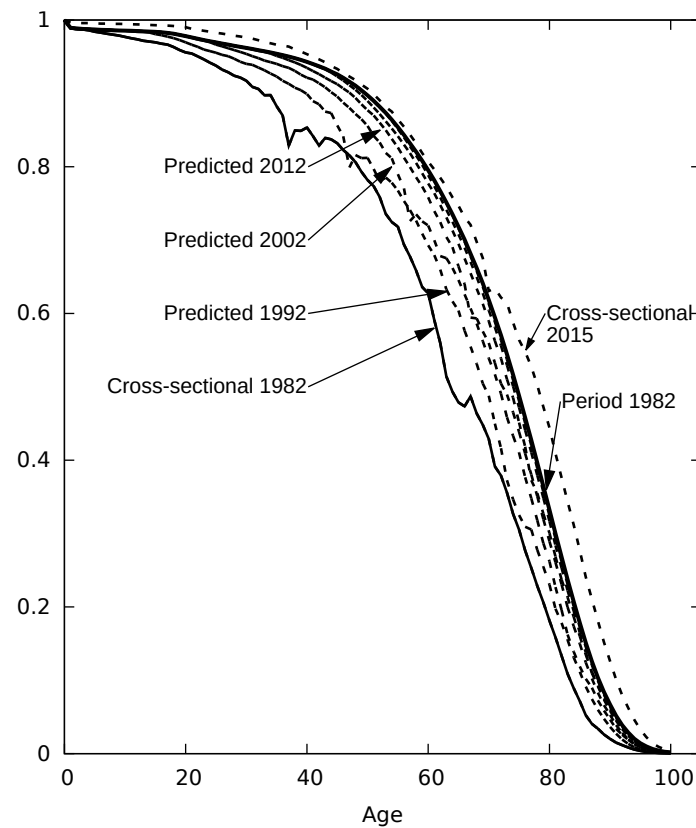


Figure 18: Forecast of the male 1982 cross-sectional prevalence of survival under the hypothesis of a constant mortality (1982) in 1992, 2002, 2012, 2022 and 2032. Comparison with the period life table of 1982 and the observed cross-sectional life prevalence of 2015.

cerns cross-sectional prevalence of survival in 2015. The 1982 mortality table was a good predictor of survivors but, obviously, mortality did not stop decreasing af-

ter 1982 and especially after 70 years where the cross-sectional is much greater than the period. For men, the “expected” life in 1982,  $e_0(1982)$ , is reached by the effective duration  $d_0$  only after 24.1 years in February 2006.

This is very similar to Fig. 2 and Fig. 5 when the economic activity of 1978 was projected in the future using the transitions measured between 1977 and 1978 up to the limit (or forward prevalence) and finally compared with the cross-sectional prevalence observed in the next censuses.

We can also say that life expectancy at birth is an advanced indicator of mortality. This advance is currently around 24 years but it overestimates the actual mortality in France of about 6 years.

Some economists, demographers and epidemiologists mixed age-specific cross-sectional prevalences of activity or disability within a period survival curve in order to summarize the resulting curve by its area and naming it the “working life expectancy” (Wolfbein, 1949) or “health expectancy” (Sullivan, 1971). These indices are hybrid, mixing a potential future of mortality decline with a observed prevalence which can’t be constant because its components the in and out flows are not constant over time in particular because the mortality is not constant and its life expectancy is different from the C.A.L.

We proposed in 1992 (Bonneuil et al., 1992) to calculate two indices, (a) a purely cross-sectional, calculated as the cross-sectional prevalence of activity or disability at each age multiplied by the cross-sectional prevalence of survival at the same age and (b) a second, a purely period index, that is to say computed only from the age-specific flows between states (healthy, disabled, dead) observed during a given period, so that their comparison will highlight the tempo effect of mortality but also of disability (Luy et al., 2018).

In this end of the chapter, we simply transposed the two demographic analysis techniques (economic status change models of 1980 and the mortality models of 1986) to the analysis of the more recent phenomenon that is the evolution of disability and dependence.

With the availability of more and more cross-longitudinal surveys, such as the Longitudinal Studies Of Aging (LSOA-I and LSOA-II) or the Health and Retirement Study (HRS), the concept of health expectancy developed by Sullivan has evolved into a more valuable indicator based solely on age-specific transition rates between health states and death (Lièvre et al., 2003).

The international REVES network (<https://reves.site.ined.fr/en/>) has been working since the mid-1980s for the development of such health indicators.

As a founding member of this network in the late 80’s, we developed with the help of Agnès Lièvre and Christopher Heathcote, a computer program called IMaCh (<http://euroves.ined.fr/imach>) that estimates Healthy Life Expectancy and age-specific forward prevalences (Lièvre et al., 2003). In today’s transversal-longitudinal surveys, people are unfortunately not interviewed at each of the passages, so that transitions between states have different exposure times

and must be estimated by a statistical model like the Interpolated Markov chain Model (Laditka and Wolf, 1998). The IMaCh software has already been used by specialized researchers in health and disability.

The software evolved since the early versions of 1999 through a first main publication in 2003. It is currently a software of 12,800 lines of C code. One of the main constraint is the time for the likelihood of the sample to be maximised and IMaCh users are invited to start with time intervals close to the mean interval between waves before decreasing the interval down to a single month which dramatically increases the number of matrix products and the time to maximum likelihood. Therefore, we derived our own method of optimisation (Brouard and Heathcote, 2015) to speed up the convergence and got grants from Intel Software in order to produce IMaCh Windows 64bits executables compiled with their fast Intel compiler.

But it is only recently that we have developed the concept of backward prevalence and that the experimental version 0.99 of IMaCh (2018) makes it possible to calculate backward prevalences by age. These backward prevalences are plotted with the cross-sectional prevalences as well as with the forward prevalences for useful comparisons.

With the general decline of mortality in the developed countries, we are usually observing a lower period prevalence of disability than the cross-sectional prevalence. This indicates that disability is currently declining or postponed. But by using the estimation provided by IMaCh on two almost identically designed cross-longitudinal surveys (LSOA I and LSOA II), we have been able to measure the changes in the transitions rates over a decade and found that the decline in the mortality of dependent people fell rapidly and contravened the general decline in dependency in the United States (Crimmins et al., 2009).

### *2.3. Some estimations of backward prevalences*

We will use two examples in order to highlight the various advantages of the calculation of backward probabilities.

#### *2.3.1. Did the 2008 economic crisis in Italy impact the Health Expectancies?*

Results from the 2013 longitudinal release of the Italian “Statistics on Income and Living Condition” survey (EU-SILC) (Giudici et al., 2017) show in Fig. 19 that the cross-sectional, forward as well as backward prevalences of disability measured by the Gali index are almost identical. It is a proof that the 2008 economic crisis in Italy hasn’t had a major impact on health expectancies. This stationarity is also seen when changing the Gali index with the self-rated health index or even the chronic diseases index.

Also we can remark that younger generations will be less disabled at very old ages.

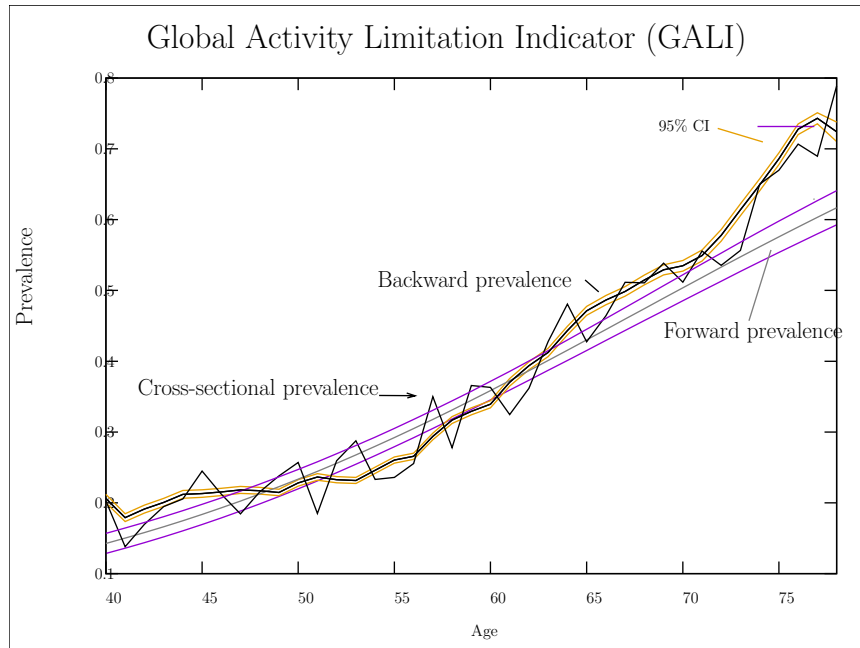


Figure 19: Italian SILC survey: backward, forward and cross-sectional prevalences are similar. 95% confidence intervals are represented. No influence of the 2008 economic crisis on health expectancy.

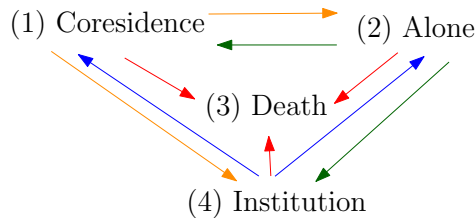


Figure 20: Diagram of the transitions between the four states.

### 2.3.2. Health and Retirement Study: changes in living arrangements

In this second example, we explored the properties of the backward prevalences using data from the recent (2017) US Health and Retirement Study with its nine waves (1998-2014). Using an earlier version of HRS (with 8 waves, we were not only interested in health changes, but also in how living conditions evolved according to the disability status of a person during his or her life cycle (Shih, 2016).

In this chapter, we are limiting our investigation to three different living arrangement statuses, “coresidence”, “alone” and “in institution” and how people move between the three states and death which is an increasingly competing risk after age 50 (see diagram 20).

- Results shown in Fig. 21 describe how living in coresidence will slightly

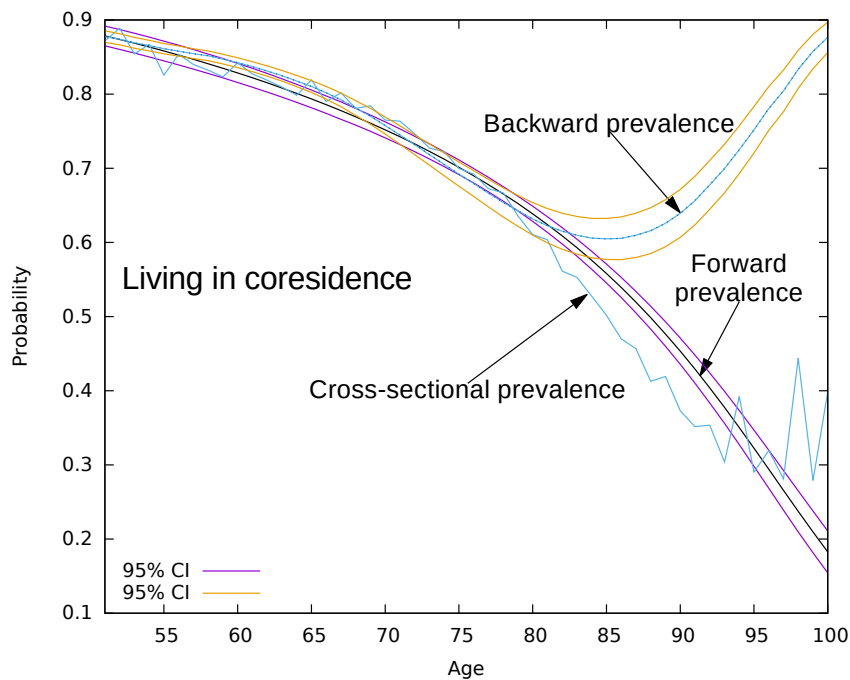


Figure 21: Cross-sectional, forward and backward prevalences with 95% confidence intervals of living in coresidence (HRS 1988-2014). Living in coresidence at old ages was more frequent in the past.

increase between age 80 to 90 (significant difference between the cross-sectional and forward prevalences at those ages). The backward prevalence also shows that for old cohorts, living in coresidence after age 85 concerned a majority of families and increased with age.

- Results in Fig. 22 indicate that “Living alone” is shifting to older ages: the peak was at age 85 (backward prevalence), is around 90 (cross-sectional) and will reach the age of 95 (forward prevalence).
- Results shown in Fig. 23 indicate that “living in an institution” will not change in the near future. It was very occasional in the past, because the backward prevalence was very low.

The most interesting change concerns the aging of the persons living alone and we can even indicate, by using projections starting from the mid period of the five waves, ie August 2003, that this modal age of 94 will be reached around 2024. (Fig. 24).

#### 2.4. Limitations concerning the estimation of forward and backward prevalences

We can't conclude this chapter without discussing the major limitations of these forward and backward prevalences. Concerning the forward prevalences, they are

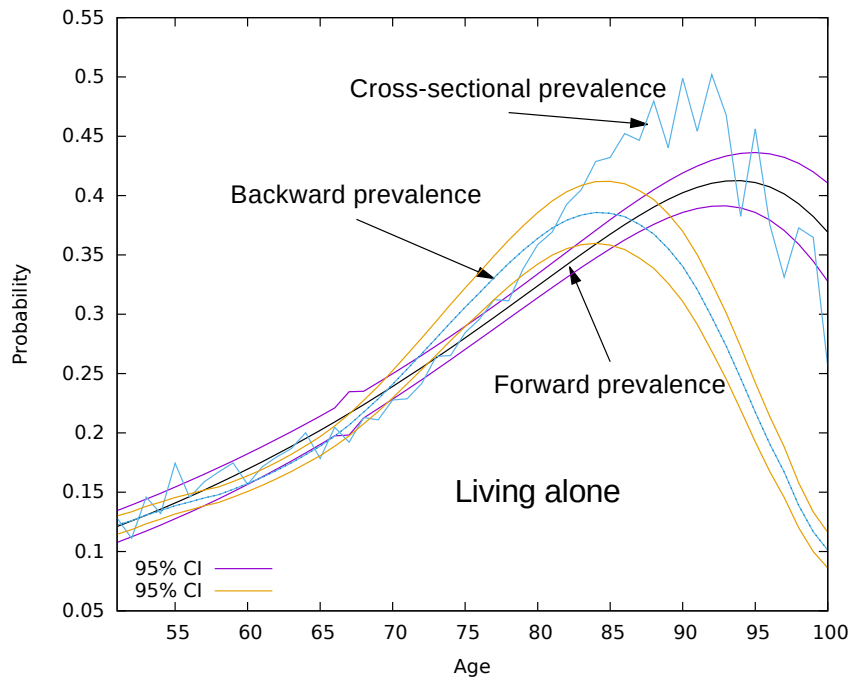


Figure 22: "Living alone": shift to older ages.

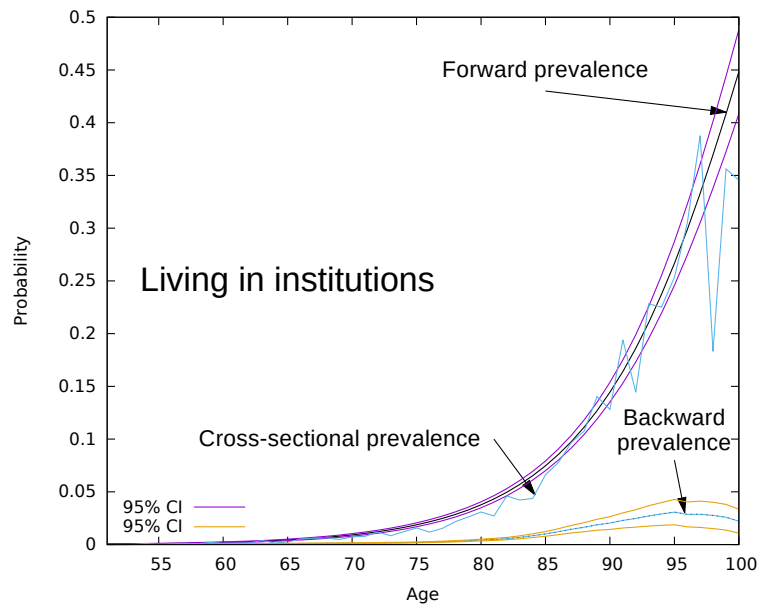


Figure 23: "Living in an institution" will not increase in the future but was very rare among the old cohorts.

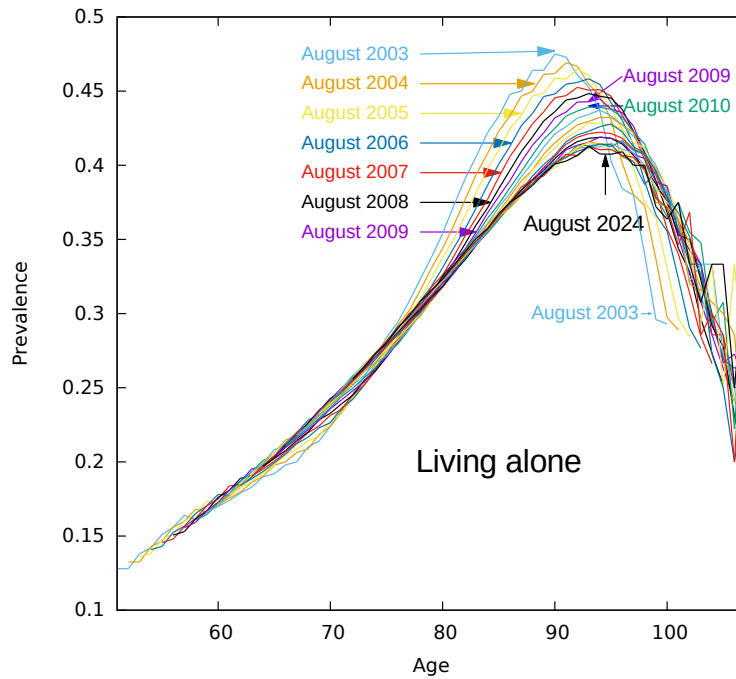


Figure 24: (Forward) Projection of prevalence of living alone from the observed in August 2003 to August 2024 (HRS 2017). The modal age will reach 94 in 2024. Results from IMaCh 0.99 .

synthesized on Fig. 25. It can be seen that from people living alone at age 50 (state 2), only half of them are still living alone ten years after at age 60 (prev(2,2)). From people living in coresidence at age 50, only 16% will be alone at age 60 (prev(1,2)). And among those living in an institution at age 50, 18% will return home alone at 60 (prev(3,2)). If these results seem plausible and interesting, we can see that those probabilities curves are not converging rapidly but at a late age, around age 90 or 95.

It is therefore difficult to obtain the estimate of the forward stationary prevalence of “Living alone” unless assumptions are made about transition forces before the age of 50. It has already been discussed in the case of the activity ratio before age 15 and was easily solved (see 1.4).

But here, there are no obvious assumptions about these transition rates before the age of 50, so that before age 90, the exact prevalence could only be framed by higher and lower prevalences. Fortunately, in the IMaCh software, the probabilities are estimated using a multinomial logistic regression model, so that it is possible, at the cost of a wider confidence interval, to extrapolate the transition probabilities outside the age range. But we can see that the 95% confidence intervals of Fig. 23 are not huge.

This discussion on the limitations does also concern the backward prevalences. It can be seen in Fig. 26 that the backward convergence is more rapid giving more



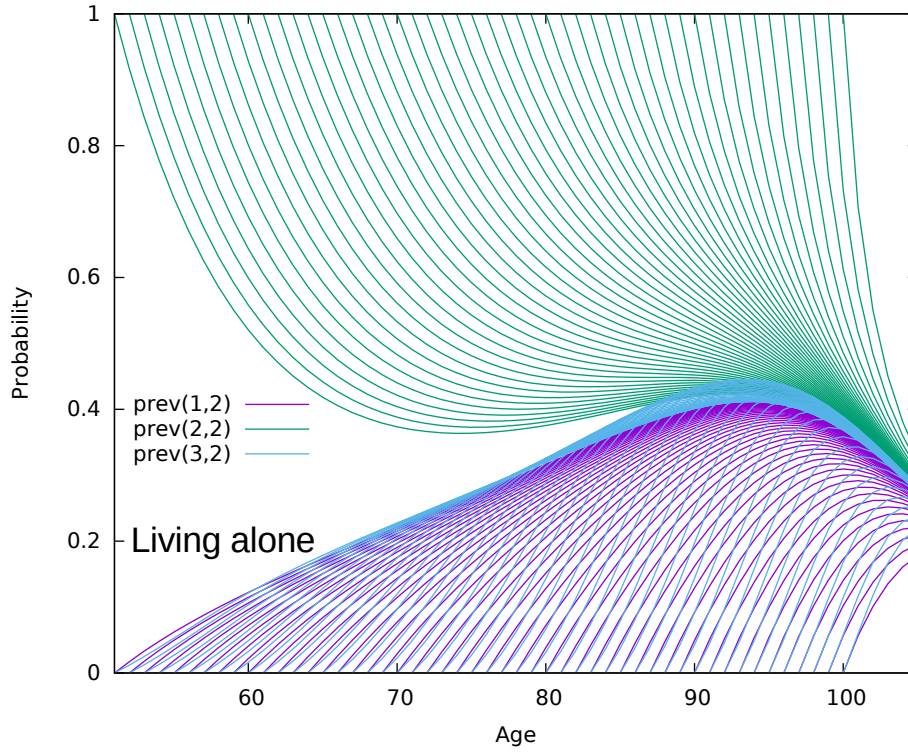


Figure 25: Slow convergence to “Living alone” (state 2): Forward probabilities to stay in the “Living alone” state (prev(2,2) or to exit from a coresidence (state 1) (prev(1,2) or from an institution (prev(3,2)).

credits to its estimation in this example.

### 2.5. Perspectives concerning the estimation of backward prevalences

Let us review first the computation of the forward prevalences in practice. IMaCh optimizes the likelihood of the sample defined by the product of the *forward* probabilities  $p_x^{ij}$ , of being observed in state  $j$  at  $x + 1$  (or next wave) being observed in state  $i$  at the first wave (at age  $x$ ) in order to (1) obtain the maximum likelihood parameters.

Then IMaCh (2) recalculates the individual transition matrices at each age, (3) builds the product of matrices using Eq. (139) until convergence is reached.

Concerning the backward prevalences, the current version 0.99 of IMaCh is simply calculating the backward matrices at each age  $x$ ,  $B_{x+1}$  using Eq. 148 which requires the estimated matrix  $P_x$  and the *observed* cross-sectional prevalences  $w_x^i$  in any state  $i$  and age  $x$  during a wave (or the average of the cross-sectional prevalences observed during different waves). We also have to ask the user to agree with a smoothing process applied to the age-specific cross-sectional prevalences. Then, IMaCh uses Eq. (156) in order to get the convergence at age  $x$  and the corresponding backward prevalences.

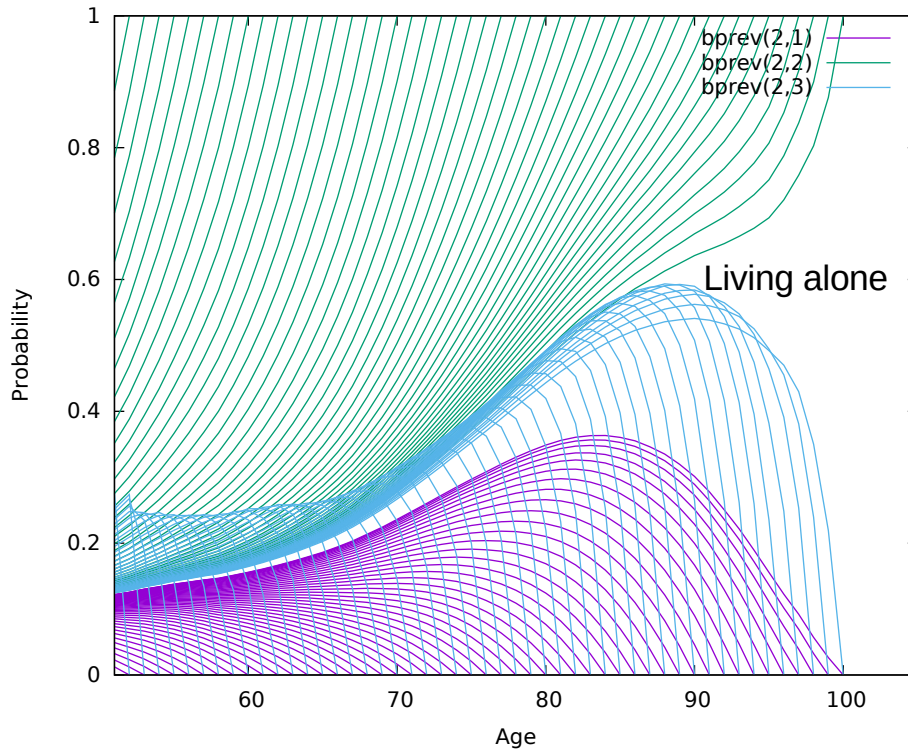


Figure 26: Convergence to the stationary backward prevalence of “Living alone” from each of the three states.

But an interesting perspective could be to build a kind of “backward” likelihood made of observed back probabilities in order to directly estimate the backward parameters and finally the backward prevalences by age.

### 3. Conclusion

Even if the concept of backprobability and backward stationary prevalence were already described in our old article of 1980 (Brouard, 1980), the mathematical theory behind this concept was never developed.

We revisited the strong and weak ergodicity properties of Markov chains and applied them to backward transitions even in the presence of an absorbing state.

Therefore we introduced a new theorem concerning the equivalence of the backward prevalences with the cross-sectional as well as with the forward prevalences in the stationary case.

In the first and in the last sections, we tried to persuade the reader that both concepts of forward and backward prevalences could be of great interest to analyze sociological changes in our modern societies.

We also know that the estimation of such prevalences requires a lot of data and computing. Data provided by cross-longitudinal surveys are more and more

numerous in the US, in Asia and in Europe. And most of the time, they are made available to researchers free of charge.

Similarly, tedious estimates of these age-transition forces as well as both backward and forward prevalences can only be made by appropriate software, such as IMaCh. Let us mention two other softwares offering multistate life tables analyses, currently limited to the calculation of health expectancies but which could be extended to the calculation of backward and forward prevalences: the SPACE (Stochastic Population Analysis for Complex Events) program, developed by Liming Cai (Cai et al., 2010) and ELECT (Estimation of life expectancies using continuous-time multi-state survival models) from Ardo van den Hout which utilizes the ‘msm’ package for R developed by Christopher H. Jackson (Jackson, 2011).

Fortunately, these two softwares, like IMaCh, are GPL licensed softwares that can be used and modified freely. Hopefully they will be used and improved in the future by new generations of economists, demographers and epidemiologists.

#### **4. Acknowledgements**

The author would like to thank a few persons: (a) Yao-Chi Shih, from the National Chung Cheng University of Taiwan for his idea to study living arrangements in the USA as a multistate model; (b) Cristina Giudici and Eleanora Trappolini from the La Sapienza University of Rome in Italy for their idea to measure if the Italian economic crisis of 2008 did affect the health expectancies by using the Italian SILC survey; (c) Saito Yasuhiko from the Nihon University of Tokyo in Japan for his continuous support to this research. Also the IMaCh software got a grant from the Japan Society for the Promotion of Science (Grant-in-Aid for Scientific Research # 25293121 and Intel software provided free licenses for the Intel C compiler, mostly efficient on Windows 64bit platforms. The author would like to thank Marília Nepomuceno, postdoctoral researcher at the Max Planck Institute for Demographic Research in Rostock, Germany, for her help to improve the readability of, otherwise, more obscure sections and an anonymous referee.

#### **5. References**

##### **References**

Bonneuil, N., Brouard, N., Robine, J.-M., 1992. Methods of calculation of health expectancy: application to the LSOA surveys (1984-86-88). In: 5th meeting of the International Network on Health Expectancy (REVES-5): Future Uses of Health Expectancy Indices, Ottawa, February 19-20-21. p. 10. <http://eurorevs.ined.fr/imach/Bonneuil-Brouard-Robine-Ottawa-1992.pdf>.

- Brouard, N., Novembre 1980. Espérance de vie active, reprises d'activité féminine : un modèle. *Revue économique* 31 (6), 1260–1287. <https://doi.org/10.2307/3501190>, (special issue on Economy and Demography).
- Brouard, N., 1986. Structure et dynamique des populations, la pyramide des années à vivre, aspect nationaux et exemples régionaux. *Espace Populations Sociétés* (2), 157–168. <https://doi.org/10.17605/OSF.IO/2UCYH> (T. Riffe in 2018 on SocArxiv, Transl.).
- Brouard, N., Heathcote, C., 2015. IMaCh is proposing to measure Health Expectancy with one index. In: Health expectancy: Is It Possible to Measure Population Health With One Index? 28th REVES meeting, Singapore, 2–4 June 2015. <http://euroreves.ined.fr/imach/singapore-Brouard-Heathcote-2015.pdf>
- Cai, L., Hayward, M., Saito, Y., Lubitz, J., Hagedorn, A., Crimmins, E., 2010. Estimation of multi-state life table functions and their variability from complex survey data using the SPACE Program. *Demog. Res.* 22 (6), 129–158. <https://doi.org/10.4054/DemRes.2010.22.6>
- Canudas-Romo, V., Guillot, M., 2015. Truncated cross-sectional average length of life: A measure for comparing the mortality history of cohorts. *Popul. Stud.* 69 (2), 147–159. <https://doi.org/10.1080/00324728.2015.1019955>. PMID: 25871610.
- Cohen, J. E., 03 1979. Ergodic theorems in demography. *Bull. Amer. Math. Soc. (N.S.)* 1 (2), 275–295. <https://projecteuclid.org:443/euclid.bams/1183544080>.
- Crimmins, E., Hayward, M. D., Hagedorn, A., Saito, Y., Brouard, N., August 2009. Change in disability-free life expectancy for americans 70 years old and older. *Demography* 46 (3), 627–646. <https://doi.org/10.1353/dem.0.0070>.
- Giudici, C., Brouard, N., Coppola, L., Trappolini, E., 2017. Disability trends among Italian adult population. A multistate approach. In: Squaring the Curve in Less Developed Countries: Thinking Public Policies for Healthy Life Expectancies. XXVIII International Population Conference, Cape-Town (South Africa) 29 Oct-4 Nov 2017, Poster.
- Goldstein, J. R., Wachter, K. W., 2006. Relationships between period and cohort life expectancy: Gaps and lags. *Popul. Stud.* 60 (3), 257–269. <https://doi.org/10.1080/00324720600895876>. PMID: 17060053.
- Guillot, M., 2003. The cross-sectional average length of life (CAL): A cross-sectional mortality measure that reflects the experience of cohorts. *Popul. Stud.* 57 (1), 41–54. <https://doi.org/10.1080/0032472032000061712>.

- Jackson, C. H., 2011. Multi-state models for panel data: The msm package for R. *J. Stat. Softw.* 38 (8), 1–29. <http://www.jstatsoft.org/v38/i08/>.
- Laditka, S. B., Wolf, D. A., 1998. New methods for analyzing active life expectancy. *J. Aging Health* 10 (2), 214–241. <https://doi.org/10.1177/089826439801000206>.
- Le Bras, H., 1971. Éléments pour une théorie des populations instables. *Population* 26 (3), 525–572. <https://doi.org/10.2307/1529142>.
- Lièvre, A., Brouard, N., Heathcote, C., 2003. The Estimation of Health Expectancies from Cross-longitudinal studies. *Math. Popul. Stud.* 10 (4), 221–248. <https://doi.org/10.1080/713644739>.
- Lopez, A., 1961. Problems in stable population theory. Office of population research, Princeton University.
- Luy, M., 2006. Mortality tempo-adjustment: An empirical application. *Demog. Res.* 15 (21), 561–590. <https://doi.org/10.4054/DemRes.2006.15.21>.
- Luy, M., Sauerberg, M., et al., 2018. Could the actual decrease in life expectancy be caused by “Tempo Effects”? In: REVES@30: Past, Present, and Future Trends in Population Health. <https://globalagingandcommunity.com/reves-2018/program/>.
- Shih, Y.-C., 2016. Living alone and subsequent living arrangement transitions among older Americans. Ph.D. thesis, University of Massachusetts Boston, Boston, MA.
- Sullivan, D. F., 1971. A single index of mortality and morbidity. *HSMHA health reports* 86 (4), 347–54.
- Wolfbein, S. L., 1949. The Length of Working Life. *Population Studies* 3 (3), 286–294.