Contents

1 A multi-state transition model

The transition model is based on the method developpped by Laditka and Wolf [1]. In theory, there is an unlimited number of states, but in a practical way, it is reasonnable to have less than 4 or 5 states, when the number of transitions is small. In this manual we introduce a 3-state transition model as an example. Instead of using real data, like data from the Longitudinal Study on Aging which can be directly obtained from the NIH, we used simulated data.

Thus our example data set consists in 8200 individuals aged 70+ interviewed in 1984 (mainly in July) and reinterviewed in 1986, 1988 and 1990.

As shown on figure 1, we coded 1 for the disability-free state, 2 for disabled state and 3 if the individual is died.

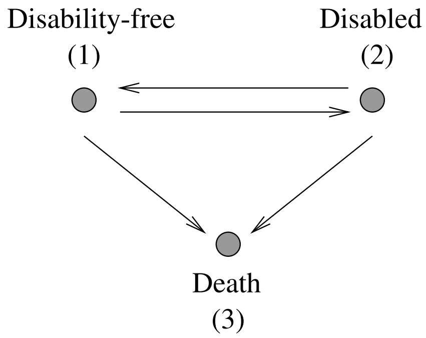

Figure 1: States and possible transitions across states

The data set consists of a data survey with at least 2 waves but no more than 4. Using the maximum likelihood estimation method, IMACH estimates the parameters of the transition probabilities and the covariance matrix of these parameters. Introducing a Markov chain on an interpolation grid, we are able to estimate transition probabilities where the time (or age) unit can be a month, a quarter or a year. Finer is the grid, better are the results.

Figure 2 exhibits the monthly histrogram of 8000 individuals at each interview. We simulated three waves whose duration varies from 23 months to 74 months with a mean of 25 months The spacing time of 74 months is the case of an individual interviewed in 1984, lost in 1986 and 1988 and finally reinterviewed in 1990. Some people died during the months following the interview up to 4 months after the interview for the earlier. This variable delay denotes the importance in taking in account the individual duration of the waves.

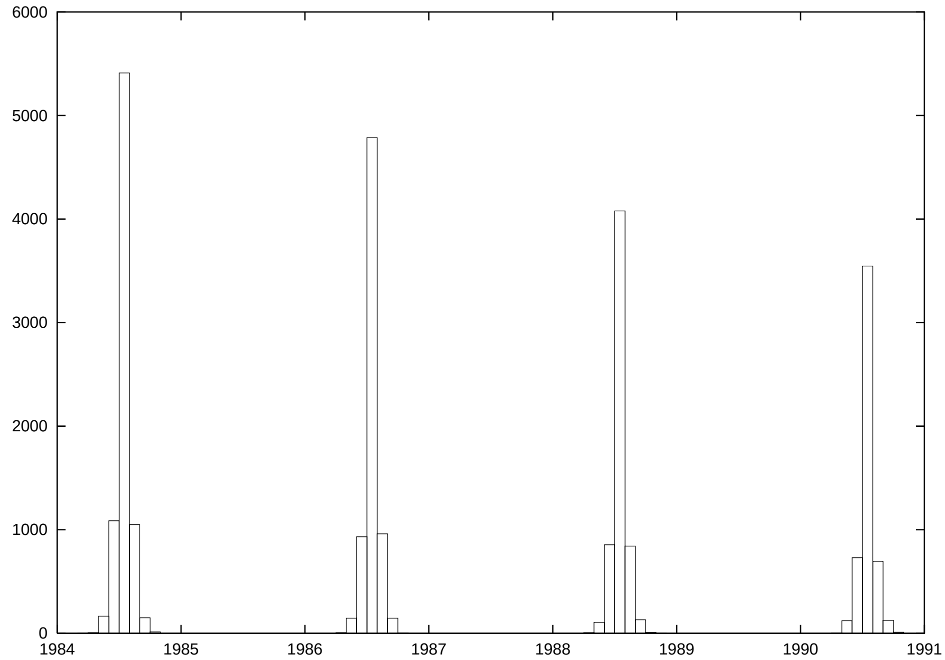

Figure 2: Monthly distribution of two to four interviews of 8200 simulated from January 1984 to December 1990

2 Mathematical background

2.1 Estimating transition probabilities

The Markov chain we use to model transitions between health states has two non-absorbing states, coded 1 for disability-free and 2 for disabled, and one absorbing state of Death which is coded 3. Let $X(x)$ denotes the state of an individual aged x at time t. After time h this individual is supposed to be in state $X(x+h)$. Assume that $X(x)$ is a non-homogeneous discrete parameter Markov chain on these three states with transition probabilities:

$$ {}_hp_x^{jk} = Pr(X(x+h) = k | X(x) = j)\,, $$

and one step transition matrix:

$$ {}_hP_x = \left({}_hp_x^{jk}\right) = \begin{pmatrix} {}_hp_x^{11} & {}_hp_x^{12} & {}_hp_x^{13} \\ {}_hp_x^{21} & {}_hp_x^{22} & {}_hp_x^{23} \\ 0 & 0 & 1 \end{pmatrix}\,. \tag{1} $$

A natural parametrisation for the transition probabilities is the multinomial logit. This parametrisation implies that the dependence on age x is of the relatively simple form of linear log(partial odds). If we assumed that h is fixed (e.g. a month or trimester) the model is

$$\log \frac{{}_h p_x^{jk}}{{}_h p_x^{jj}} = \alpha_{jk}(h) + \beta_{jk}(h)x, \quad j \neq k. \tag{2}$$

A more detailed treatment with commensurate increase in computational effort would involve the examination of a more complex dependence on age, and/or cohort, and the introduction of covariates.

Let $t_1, t_2, ...$ be the times of scheduled follow-up of the study as shown on Figure 3.

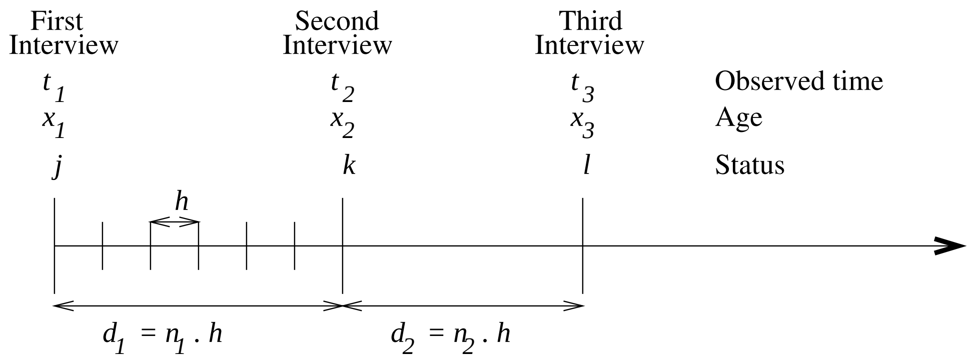

Figure 3: Follow-up of an individual i in state j aged x_1 at first interview (time t_1) reinterviewed at times t_2 and t_3.

Let us suppose that individual i observed at state j at first examination (time t_1 and age x_1), is classified in state k and age $x_2 = x_1 + d_1$ at second interview $(t_2 = t_1 + d_1)$.

It is assumed that the lag between two interviews is expressed in terms of elementary time units (h), i.e. $d_1 = n_1 h$.

Then the contribution so far of individual i to the likelihood is:

$$L^{(i)}_{t_1 - t_2} = {}_{d_1} p^{jk}_{x_1},$$

where the probability on the right hand side is the (j,k)th element of the matrix product:

$$\prod_{u=1}^{n_1} {}_h P_{x_1+(u-1)h} = \left({}_{d_1} p^{jk}_{x_1}\right) . \tag{3}$$

If this individual is observed only once more at say time $t_3 = t_2 + d_2$ ($d_2 = n_2 . h$) and noted to be in state l, then its further contribution to the likelihood is ${}_{d_2}p_{x_2}^{kl}$ $(x_2 = x_1 + d_1)$. In this case the component of the total likelihood due to individual i is

$$L^{(i)} = \left({}_{d_1}p_{x_1}^{jk}\right) \times \left({}_{d_2}p_{x_2}^{kl}\right).$$

Similarly, contributions will come from all members of the sample. One observes that the formation of the likelihood is no trivial matter since there is no simple analytical expression for the higher order transition probabilities in (3).

If an individual is not alive at the time of the interview, the date of death has to be known in order to calculate the contribution to the likelihood. If this date is unknown, we assumed that the individual died at the mid interval. If it is alive but its status is unknown, there is no contribution until the next interview.

In this way, the log-likelihood of a sample of N individuals can be formed:

$$l = \sum_{i=1}^{N} \log L^{(i)} . \tag{4}$$

The parameters $\alpha_{ij}(h)$ and $\beta_{ij}(h)$ of (2) are estimated by maximising this quantity.

If $\underset{\sim}{\theta}$ denotes the vector of parameters and $\hat{\underset{\sim}{\theta}}$ its maximum likelihood estimator, then it follows from standard theory that for large sample size N the MLE $\hat{\underset{\sim}{\theta}}$ is approximatively normally distributed with mean $\underset{\sim}{\theta}$ and covariance matrix

$$\lim_{N\to\infty} \mathrm{E}(\hat{\underset{\sim}{\theta}}) = \underset{\sim}{\theta} ,$$
$$V(\underset{\sim}{\theta}) = \frac{1}{N} I^{-1}(\underset{\sim}{\theta}) , \tag{5}$$

where $I(\underset{\sim}{\theta})$ is the information matrix computed at the true value $\underset{\sim}{\theta}$. This implies the asymptotic normality of the transition probabilities and health expectancies, as now discussed.

2.2 Estimation of health expectancies and their variances

For convenience take the time/age unit h=1 so that in what follows the temporal unit is suppressed in the notation. Recall, however, that different values of h will give different numerical estimates of the parameters. Given implementation of the maximum likehood procedure described above the health expectancies are estimated as follows.

Let e_x^{ij} be the expected subsequent time spent in state j by an individual who was in state i at age x. For such individuals, e_x^{ij} is the health expectancy of state j. If

$${}_yI_x^{ij} = \begin{cases} 1 & \text{if } X(x+y) = j \text{ given } X(x) = i \, , \\ 0 & \text{otherwise.} \end{cases}$$

then

$$O_x^{ij} = \sum_{y=1}^{\infty} {}_yI_x^{ij}$$

is the subsequent time spent in state j by individuals in state i at age x. Now the mean of ${}_yI_x^{ij}$ is

$$\mathrm{E}({}_yI_x^{ij}) = \Pr(X(x+y) = j|X(x) = i) = {}_yp_x^{ij}$$

and hence

$$e_x^{ij} = \mathrm{E}(O_x^{ij}) = \sum_{y=1}^{\infty} {}_yp_x^{ij} \tag{6}$$

Thus knowledge of the transition probabilities ${}_yp_x^{ij}$ permits calculation of the health expectancies e_x^{ij} .

In fact the probabilities ${}_yp_x^{ij}(\theta)$ are estimated by ${}_yp_x^{ij}(\hat{\theta})$ where $\hat{\theta}$ is the maximum likelihood estimator of the previous section. Hence e_x^{ij} is estimated by

$$\widehat{e_x^{ij}} = e_x^{ij}(\hat{\theta}) = \sum_{y=1}^{\infty} {}_yp_x^{ij}(\hat{\theta}) \tag{7}$$

Clearly $e_x^{ij}(\hat{\theta})$ is a consistent normally distributed estimator of e_x^{ij}.
The Taylor approximation

$${}_yp_x^{ij}(\hat{\theta}) \simeq {}_yp_x^{ij}(\theta) + (\hat{\theta} - \theta)' \frac{\partial_y p_x^{ij}(\theta)}{\partial \theta}.$$

implies the large sample variances and covariances

$$\mathrm{Var}(e_x^{ij}(\hat{\theta})) = \sum_y \sum_u (\frac{\partial}{\partial \theta} {}_yp_x^{ij}(\theta))' V(\theta) \frac{\partial}{\partial \theta} {}_up_x^{ij}(\theta),$$
$$Cov(e_x^{ij}(\hat{\theta}), e_x^{ik}(\hat{\theta}) = \sum_y \sum_u (\frac{\partial}{\partial \theta} {}_yp_x^{ij}(\theta))' V(\theta) \frac{\partial}{\partial \theta} {}_up_x^{ik}(\theta),$$

where $V(\theta)$ is the covariance matrix (5).

If one were interested in the health expectancy

$${}_ye_x^{ij} = \sum_{u=1}^{y} {}_up_x^{ij}$$

over the interval $(x, x+y)$, given the initial state i at age x, then the above formulae hold modified by the upper limit y in the summations.

Health expectancies irrespective of the initial state are weighted averages of e_x^{1j} and e_x^{2j}. The weights are the proportions π_x^1, π_x^2 of the population aged x that are in states 1 and 2 respectively. Thus

$$e_x^j = \pi_x^1 e_x^{1j} + \pi_x^2 e_x^{2j} \tag{8}$$

is the (marginal) health expectancy of state j at age x. Its estimator $e_x^j(\hat{\theta})$ is the same linear combination of $e_x^{ij}(\hat{\theta})$, i=(1,2), and as such is asymptotically normally distributed with mean e_x^j and variance

$$\mathrm{Var}(e_x^j(\hat{\theta})) = \sum_{i=1}^{2}(\pi_x^i)^2 \,\mathrm{Var} e_x^{ij}(\hat{\theta}) + 2\pi_x^1\pi_x^2 Cov(e_x^{1j}(\hat{\theta}), e_x^{2j}(\hat{\theta})).$$

In fact $e_x^1(\hat{\theta})$ and $e_x^2(\hat{\theta})$ are asymptotically normally distributed with covariance

$$\begin{aligned} Cov(e_x^1(\hat{\theta}), e_x^2(\hat{\theta})) &= (\pi_x^i)^2 Cov(e_x^{11}(\hat{\theta}), e_x^{12}(\hat{\theta})) \\ &+ 2\pi_x^1\pi_x^2 \left[Cov(e_x^{11}(\hat{\theta}), e_x^{22}(\hat{\theta})) + Cov(e_x^{12}(\hat{\theta}), e_x^{21}(\hat{\theta}))\right] \\ &+ (\pi_x^2)^2 Cov(e_x^{21}(\hat{\theta}), e_x^{22}(\hat{\theta})). \end{aligned}$$

These formulae apply if π_x^1 and π_x^2 are known. However, this will generally not be the case and the proportions will have to be estimated from the sample. One way of proceeding is to use the age specific proportions. Recall that these are obtained from population surveys and are commonly used to estimate health expectancies from cross-sectional data. We have chosen not to use observed prevalences but rather what we called quasi-stationary prevalences. They may be of independent interest and are discussed in the next section.

2.3 Stationary prevalences

Consider inhomogeneous products of non-negative stochastic matrices obtained from a markov process defined on 2 non-absorbing states. We are interested in the behaviour before absorption of the transition matrices ${}_tP_{x-t}$ as $t \to \infty$.

Let

$${}_tw(x) = \begin{pmatrix} {}_tw^{\cdot 1}(x) \\ {}_tw^{\cdot 2}(x) \end{pmatrix}.$$

be the conditional distribution at age x. The quantity ${}_tw^{\cdot 1}(x)$ is the proportion of individuals in state 1 at age x and can be expressed in terms of ${}_tp_{x-t}^{i1}$ and ${}_tp_{x-t}^{i2}$ ($i = 1, 2$) that are the probabilities to be, respectively, in state 1 and 2 at age x. It follows that

$$\begin{aligned} {}_tw^{\cdot 1}(x) &= \frac{{}_tp_{x-t}^{i1}}{{}_tp_{x-t}^{i1} + {}_tp_{x-t}^{i2}} \\ {}_tw^{\cdot 2}(x) &= \frac{{}_tp_{x-t}^{i2}}{{}_tp_{x-t}^{i1} + {}_tp_{x-t}^{i2}}. \end{aligned}$$

We call ${}_tQ_{x-t}$ a submatrix (2 x 2) of the transition matrix product ${}_tP_{x-t}$ defined in (3).

$$ {}_tQ_{x-t} = \begin{pmatrix} {}_tp_{x-t}^{11} & {}_tp_{x-t}^{12} \\ {}_tp_{x-t}^{21} & {}_tp_{x-t}^{22} \end{pmatrix} . $$

If ${}_tw(x) = w(x)$ for all $t \geq 0$ then $w(x)$ is a stationary conditional distribution. Given times t_1 and t_2, we can deduce from the process that

$$ {}_{t_1+t_2}p_{x-t_1-t_2}^{ij} = {}_{t_1}p_{x-t_1}^{ij} \times {}_{t_2}p_{x-t_1-t_2}^{ij} $$

where t_1, $t_2 \geq 0$ and $j = 1, 2$.

From this formula, we can make the assumption that a function λ depending of t_1,t_2 and x exists and satisfy

$$ {}_{t_1}w'(x) \times {}_{t_2}Q_{x-t_1-t_2} = \lambda(t_1, t_2, x) \times {}_{t_1+t_2}w'(x) $$

If such a $w(x)$ exists as t_1, $t_2 \to \infty$, $w(x)$ should verify

$$ w'(x)\ Q(x) = \lambda(x)\ w'(x) $$

Since $Q(x)$ is a non-negative substochastic matrix, it follows that $Q(x)$ is irreductible. Because $w(x)$ is non-negative, we can assert, according to the theorem of Perron-Frobenius (Theorem 3.4 in Seneta [2]), that $\lambda(x)$ is the maximum eigenvalue of $Q(x)$, λ_1 say, and that $w(x)$= $v(x)$ where $v(x)$ is the left eigenvector (normalised to make $v^1(x) + v^2(x) = 1$) of $Q(x)$ corresponding to λ_1.

Then the stationary prevalence of disability at age x is defined to be

$$ w^2(x) = \lim_{t\to\infty} {}_tw^{i2}(x), \tag{9} $$

provided the limit exists independently of i. From a demographic point of view the limit in (9) is a useful fiction (human ages cannot be negative) and it should be interpreted as

$$ {}_tw^{i2}(x) = w^1(x) + o(1) $$

where the $o(1)$ term becomes negligible even for moderate values of t and positive $x - t$.

2.4 Health expectancies estimated with the stationary prevalences

A consequence of these considerations is that the (marginal) health expectancy of state j at age x, given only that the individual is alive at x, is as in (8) with the prevalence π_x^2 replaced by $w^2(x)$. Writing $w^2(x, \underset{\sim}{\theta}) = w^2(x)$

to indicate the parametric dependence of this quasi-stationary prevalence, we have for j=(1,2),

$$e_x^j(\theta) = \left[1 - w^2(x,\underset{\sim}{\theta})\right] e_x^{1j}(\theta) + w^2(x,\underset{\sim}{\theta}) e_x^{2j}(\theta) \tag{10}$$

It is these health expectancies that we take as the marginal expectancies of interest.

The estimate $e_x^j(\hat{\theta})$ of $e_x^j(\theta)$ is obtained by replacing $\underset{\sim}{\theta}$ with $\underset{\sim}{\hat{\theta}}$ in (10). A realisation of $e_x^j(\hat{\theta})$ can be calculated without difficulty once the maximum likelihood estimation of $\underset{\sim}{\theta}$ has been done but its variance must be estimated indirectly since $w^2(x,\underset{\sim}{\theta})$ is not known explicitly. Proceed as follows. Let

$$e_x^j(i,t,\theta) = \left[1 -_t w^{i2}(x,\underset{\sim}{\theta})\right] e_x^{1j}(\theta) +_t w^{i2}(x,\underset{\sim}{\theta}) e_x^{2j}(\theta)$$

the quantity whose limit as $t \to \infty$ is $e_x^j(\theta)$ of (10). With $e_x^j(i,t,\theta)$ its estimator we have, using the definitions of the $e_x^{ij}(\theta)$ and rearranging,

$$\begin{aligned} e_x^j(i,t,\hat{\theta}) - e_x^j(i,t,\theta) &= \sum_{y=0}^{\infty} {}_y p^{1j}{}_x(\hat{\theta}) - {}_y p^{1j}{}_x(\theta) \\ &+ {}_t w^{i2}(x,\hat{\theta}) \left[{}_y p_x^{2j}(\hat{\theta}) -_y p_x^{1j}(\hat{\theta})\right] \\ &- {}_t w^{i2}(x,\theta) \left[{}_y p_x^{2j}(\theta) -_y p_x^{1j}(\theta)\right] \end{aligned}$$

Now use the Taylor approximation for ${}_y p_x^{ij}(\hat{\theta})$ as before and also that for ${}_t w^{i2}(x,\hat{\theta})$,

$${}_t w^{i2}(x,\hat{\theta}) \simeq {}_t w^{i2}(x,\theta) + (\hat{\theta}-\theta)' \frac{\partial}{\partial\theta}{}_t w^{i2}(x,\theta).$$

Substitution of these approximations and simplification yields

$$e_x^j(i,t,\hat{\theta}) - e_x^j(i,t,\theta) \simeq \sum_{y=0}^{\infty} (\hat{\theta}-\theta)' \frac{\partial}{\partial\theta} g(i,t,y,x,j,\theta),$$

where

$$g(i,t,y,x,j,\theta) =_y p_x^{1j}(\theta) + {}_t w^{i2}(x,\theta) \left[{}_y p_x^{2j}(\theta) -_y p_x^{1j}(\theta)\right].$$

Assuming uniform differentiability with respect to θ and proceeding to the limit as $t \to \infty$, gives the key result

$$e_x^j(\hat{\theta}) - e_x^j(\theta) \simeq \sum_{y=0}^{\infty}(\hat{\theta} - \theta)' \frac{\partial}{\partial\theta} g(y, x, j, \theta), \tag{11}$$

with

$$g(y, x, j, \theta) =_y p_x^{1j}(\theta) + w^2(x, \theta) \left[_y p_x^{2j}(\theta) -_y p_x^{1j}(\theta) \right].$$

Assymptotic normality of $e_x^j(\hat{\theta})$ is an immediate consequence of (11). The second moments are

$$\mathrm{Var}(e_j(x, \hat{\theta})) \simeq \sum_{u=0} \sum_{y=0} \left(\frac{\partial}{\partial \underset{\sim}{\theta}} g(u, x, j, \underset{\sim}{\theta}) \right)' V(\underset{\sim}{\theta}) \left(\frac{\partial}{\partial \underset{\sim}{\theta}} g(u, x, j, \underset{\sim}{\theta}) \right), \tag{12}$$

$$Cov(e_j(x, \hat{\theta}), e_k(x, \hat{\theta})) \simeq \sum_{u=0} \sum_{y=0} \left(\frac{\partial}{\partial \underset{\sim}{\theta}} g(u, x, j, \underset{\sim}{\theta}) \right)' V(\underset{\sim}{\theta}) \left(\frac{\partial}{\partial \underset{\sim}{\theta}} g(u, x, k, \underset{\sim}{\theta}) \right).$$

The reason for this circuitous route to (12) is that the derivative $\frac{\partial}{\partial\theta} w^2(x, \theta)$ cannot be calculated directly. On the other hand calculating $\frac{\partial}{\partial\theta} {}_t w^{i2}(x, \theta)$ presents no problems once the derivatives of the transition probabilities are known. Calculation of these latter derivatives is a precondition for estimating second moments and was done numerically for the data set after the maximum likelihood estimator $\hat{\theta}$ had been found. Thus the evaluation of (12) involves calculating the realised value of the derivative of ${}_t w^{i2}$ for large t, meaning of the order of 200 months for our data set.

References

[1] Laditka S. and Wolf D. New methods for analysing active life expectancy. *Journal of Aging and Health*, 1997.

[2] Seneta E. *Non-negative matrices.* John Wiley and Sons, New York, 1973.