Empirical Transition Matrix of Multi-State Models: The etm Package

Arthur Allignol
University of Freiburg

Martin Schumacher
University Medical Center, Freiburg

Jan Beyersmann
University of Freiburg

Abstract

Multi-State models provide a relevant framework for modelling complex event histories. Quantities of interest are the transition probabilities that can be estimated by the empirical transition matrix, that is also referred to as the Aalen-Johansen estimator. In this paper, we present the R package **etm** that computes and displays the transition probabilities. **etm** also features a Greenwood-type estimator of the covariance matrix. The use of the package is illustrated through a prominent example in bone marrow transplant for leukaemia patients.

Keywords: Aalen-Johansen estimator, competing risks, covariance matrix, current leukaemia free survival, R.

1. Introduction

When dealing with complex event history data in which individuals may experience more than one single event type, multi-state models provide a relevant modelling framework. A multi-state model is a stochastic process that at any time occupies one of a set of discrete states, which can be in biomedical applications health conditions (e.g., healthy, diseased, dead) or disease stages (e.g., cancer grades or viral loads in AIDS studies). Data typically consist of times of occurrence of transitions between the states and the types of transitions that occur. Observation of the multi-state process is usually subject to right-censoring and/or left-truncation. A helpful way to approach multi-state models is through drawings, with boxes representing the states and arrows the possible transitions. For instance, the usual survival situation may be depicted in the multi-state framework as in Figure 1. Here patients enter the study in state 0 (e.g., alive without any disease) and can reach state 1 (e.g., death). State 1 is said to be absorbing, as no further transitions are modelled. States that are not absorbing

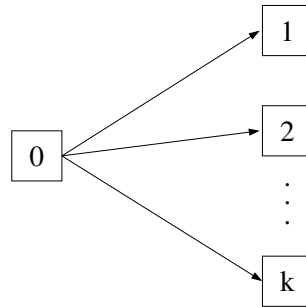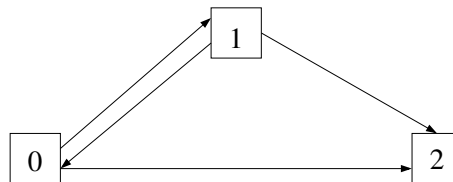Figure 1: Representation of the usual survival data situation.

Figure 2: A competing risks model with k failure types.

Figure 3: Illness-death model with recovery.

are said to be transient.

A first extension of the univariate survival data is the competing risks situation illustrated in Figure 2. This model has a transient state 0 and k absorbing states that could represent several possible causes of death. In general, the competing event states will be different types of some first event (Putter *et al.* 2007). Note that the present process description of competing risks does not impose an 'independent risks' assumption (Moon *et al.* 2002).

A well-known and more complex multi-state model is the illness-death model with recovery (Figure 3). This model was used primarily for studying transitions between being healthy (state 0), diseased (state 1) and dead (state 2). In this model, the disease is curable as transitions between the states 0 and 1 are possible in both directions. More generally, such a model can be used to study the effect of any discrete, reversible time-dependent covariate on the transition toward the ultimate state.

More examples of multi-state models can be found in Andersen *et al.* (1993); Hougaard (2000); Andersen and Keiding (2002); Putter *et al.* (2007); Andersen and Pohar Perme (2008), and in Klein and Shu (2002) with a special focus on bone marrow transplantation examples.

Usually, a multi-state process is assumed to be time-inhomogeneous Markov. I.e., the future development of the process only depends on the current state and the time elapsed since time origin. The stochastic behaviour of such a model is completely described through the transition hazards. Of interest are also the transition probabilities. For example, the probability of

a transition from state "alive" to "death" in the survival data situation displayed in Figure 1 may be estimated by one minus the Kaplan-Meier estimator. The Kaplan-Meier estimator has been generalised to arbitrary Markov processes with a finite number of states by Aalen and Johansen (1978), see also Aalen (1978), and Fleming (1978a,b) for related work. The estimator of the transition probability matrix is either denoted as the empirical transition matrix or the Aalen-Johansen estimator.

In R (R Development Core Team 2010), the **survival** package (Therneau and Lumley 2010) permits to compute the Kaplan-Meier estimator, while package **cmprsk** (Gray 2010) estimates transition probabilities in competing risks (also denoted cumulative incidence functions). The **survival** package allows for left-truncated data while **cmprsk** does not. Outputs of these functions can be used to compute the empirical transition matrix in some multi-state models like the illness-death model without recovery for which the elements of the matrix take an explicit form. However this is not generally the case. The **changeLOS** package (Wangler *et al.* 2006) permits to compute the Aalen-Johansen estimator for arbitrary multi-state models, but lacks variance estimation and cannot handle left-truncated data. Transition hazards in multi-state models, potentially subject to left-truncation and right-censoring, can be estimated in R using the **mvna** package (Allignol *et al.* 2008). The empirical transition matrix could be computed by hand as a matrix-valued function of the value returned by **mvna**, but estimation of the covariance matrix would still be lacking. These drawbacks are settled in the R package **etm** available from the Comprehensive R Archive Network (CRAN) at `http://CRAN.R-project.org/package=etm`. **etm** computes the empirical transition matrix for any finite-state multi-state models, along with the covariance estimator described in Andersen *et al.* (1993, Equation 4.4.17). The package handles both left-truncated and right-censored data.

This paper proceeds as follow: Section 2 provides statistical details on estimation of the empirical transition matrix and the covariance matrix. Section 3 describes the **etm** package and illustrates its use with an example on bone marrow transplant for patients suffering from leukaemia. A discussion is in Section 4.

2. The multi-state model framework

2.1. Data and notation

Consider a stochastic process $(X_t)_{t \in [0, +\infty)}$ with a finite state space $\mathcal{S} = \{0, \ldots, K\}$ and right continuous sample paths. We further suppose that it is time-inhomogeneous Markov. As usual with survival data, individuals are generally followed over a certain period of time, providing right-censored and/or left-truncated observations. We further assume that truncation and censoring is independent (Andersen *et al.* 1993, Chapter III). Simple examples are random left-truncation and random right-censoring. Also note that, e.g., the assumption of independent censoring allows the censoring process to depend on the currently occupied state (Aalen and Johansen 1978). The transition hazard from state $i \in \mathcal{S}$ to state $j \in \mathcal{S}$, $i \neq j$ is defined as

$$\alpha_{ij}(t)dt = P(X_{t+dt} = j | X_t = i),$$

where we write dt both for the length of the infinitesimal interval $[t, t + dt)$ and the interval itself. The cumulative transition hazards are then defined as $A_{ij}(t) = \int_0^t \alpha_{ij}(u)du$, $A_{ii}(t) =$

$-\sum_{j\neq i} A_{ij}(t)$. Let

$$P_{ij}(s,t) = P(X_t = j | X_s = i), \text{ for } i, j \in \mathcal{S}, \ s \leq t,$$

be the probability that an individual who is in state i at time s will have moved to state j at a later time t, and we write $\mathbf{P}(s,t)$ for the $(K+1) \times (K+1)$ matrix of those transition probabilities. $\mathbf{P}(s,t)$ can be recovered from the transition hazards through product integration (Aalen and Johansen 1978; Gill and Johansen 1990):

$$\mathbf{P}(s,t) = \prod_{(s,t]} (\mathbf{I} + d\mathbf{A}(u)),$$

where $\mathbf{A}(t)$ is the matrix of cumulative hazards.

2.2. The empirical transition matrix

Let $N_{ij}(t)$ be the number of observed direct transitions from state i to state j up to time t and $Y_i(t)$ be the number of individuals under observation in state i just before time t.

The non-diagonal entries of the matrix of cumulative transition hazards $\mathbf{A}(t)$ may be estimated by the Nelson-Aalen estimator (Andersen *et al.* 1993; Allignol *et al.* 2008)

$$\hat{A}_{ij}(t) = \int_0^t \frac{dN_{ij}(u)}{Y_i(u)}, \ i \neq j,$$

and $\hat{A}_{ii}(t) = -\sum_{j\neq i} \hat{A}_{ij}(t)$. Then we plug in the estimator of the cumulative transition hazards matrix, leading to

$$\hat{\mathbf{P}}(s,t) = \prod_{(s,t]} (\mathbf{I} + d\hat{\mathbf{A}}(u)).$$

As $\hat{\mathbf{A}}(t)$ is a matrix of step-functions with a finite number of jumps on $(s,t]$, the product integral can be written as a finite matrix product

$$\hat{\mathbf{P}}(s,t) = \prod_{s < t_k \leq t} \left(\mathbf{I} + \Delta \hat{\mathbf{A}}(t_k) \right),$$

where the product is taken over all observed transition times in $(s,t]$. Note that the non-diagonal entries (i,j), $i \neq j$, of $I + \Delta \hat{A}_{t_k}$ are the number of observed direct $i \rightarrow j$ transitions, divided by the number of individuals under observation in state i just prior time t_k. The diagonal entries of $I + \Delta \hat{A}_{t_k}$ are such that each row equals 1.

2.3. Covariance estimation

The covariance of the empirical transition matrix may be estimated by

$$\widehat{\mathrm{cov}}(\hat{\mathbf{P}}(s,t)) = \int_s^t \hat{\mathbf{P}}(u,t)^\top \otimes \hat{\mathbf{P}}(s,u-)\widehat{\mathrm{cov}}(d\hat{\mathbf{A}}(u))\hat{\mathbf{P}}(u,t) \otimes \hat{\mathbf{P}}(s,u-)^\top, \tag{1}$$

(Andersen *et al.* 1993, Equation 4.4.17), where $\top$ denotes vector transpose and $\otimes$ the Kronecker product. This formulation enables integrated cumulative hazards to not be necessarily

continuous, leading to an estimator of the covariance matrix of the Greenwood type. That is, (1) reduces to the Greenwood estimator for survival data (Figure 1). Moreover, (1) accounts for tied observations. Formula (1) may be used for computational purpose. However, computation is facilitated using the recursion formula (Andersen *et al.* 1993, Equation 4.4.19)

$$
\widehat{\text{cov}}(\hat{\mathbf{P}}(s,t)) = \{(\mathbf{I} + \Delta\hat{\mathbf{A}}(t))^\top \otimes \mathbf{I}\}\widehat{\text{cov}}(\hat{\mathbf{P}}(s,t-))\{(\mathbf{I} + \Delta\hat{\mathbf{A}}(t)) \otimes \mathbf{I}\} \\
+ \{\mathbf{I} \otimes \hat{\mathbf{P}}(s,t-)\}\widehat{\text{cov}}(\Delta\hat{\mathbf{A}}(t))\{\mathbf{I} \otimes \hat{\mathbf{P}}(s,t-)^\top\},
\tag{2}
$$

where $\widehat{\text{cov}}(\Delta\hat{\mathbf{A}}(t))$ is a $(K+1)^2 \times (K+1)^2$ matrix defined element-wise by

$$
\widehat{\text{cov}}(\Delta\hat{A}_{kl}(t), \Delta\hat{A}_{mn}(t)) = \begin{cases} \{Y_k(t) - \Delta N_{k.}(t)\}\Delta N_{k.}(t)Y_k(t)^{-3}, \text{ for } k = l = m = n \\ -\{Y_k(t) - \Delta N_{k.}(t)\}\Delta N_{kn}Y_k(t)^{-3}, \text{ for } k = l = m \neq n \\ \{\delta_{ln}Y_k(t) - \Delta N_{kl}(t)\}\Delta N_{kn}(t)Y_k(t)^{-3}, \text{ for } k = m, k \neq l, k \neq n \\ 0, \text{ Otherwise,} \end{cases}
$$

where δ_{ln} is a Kronecker delta, and $N_{k.}(t)$ is the number of transitions from state k up to time t. This matrix can be partitioned in blocks of $(K+1) \times (K+1)$ matrices, where only the diagonal elements are non-zero.

The covariance matrix $\widehat{\text{cov}}(\hat{\mathbf{P}}(s,t))$ is also a $(K+1)^2 \times (K+1)^2$ matrix, whose (h,r)th block, $h, r \in \mathcal{S}$, when $\widehat{\text{cov}}(\hat{\mathbf{P}}(s,t))$ is written as a partitioned matrix, is the $(K+1) \times (K+1)$ matrix of covariances between the elements of the hth and the rth column of $\hat{\mathbf{P}}(s,t)$. Using this block notation is useful for locating a specific covariance in the entire matrix. For example, $\widehat{\text{cov}}(\hat{P}_{kl}(t), \hat{P}_{mn}(t))$ is situated in the (l,n)th block, at the kth line and the mth column of this block. The **etm** package provides the `trcov` method to easily extract a specific covariance.

3. Package description and illustration

3.1. Package description

The **etm** package contains the main function `etm` that computes the empirical transition matrix, six methods (`print`, `summary`, `plot`, `xyplot`, `trprob`, `trcov`), the function `etmprep` that helps to transform a data set in the wide format (i.e., with one row per individual) into the long format, i.e., with one row per individual and per transition (Putter *et al.* 2007), suitable for using `etm`, and two data sets. The `etm` function takes seven arguments:

- `data`: A data frame in the long format, *i.e.*, with one row per transition. The data frame must contain the following columns:

    - `id`: The individual identification number.
    - `from`: State from where a transition occurs.
    - `to`: State to which a transition occurs.
    - `time` or `entry` **and** `exit`: Time when a transition occurs. If the data are left-truncated, `entry` and `exit` times can be used, where `entry` is the entry time into a state and `exit` is the exit time from this state.

- `state.names`: A character vector giving the state names.

- `tra`: A quadratic matrix of logical specifying the possible transitions.

- `cens.name`: A character giving how censored observations are identified. In case of complete data, `cens.name` may be set to `NULL`.

- `s` and `t`: Times for computing $\hat{\mathbf{P}}(s, t)$. Only `s` is mandatory, and `t` will be the last data time if not specified. `s` can take any value between 0 and t.

- `covariance`: A logical specifying whether to compute the covariance matrix.

The `etm` function returns an object of class `etm` with the following components:

- `est`: The empirical transition matrix computed for each time in $(s, t]$. This is a 3-dimensional array whose first dimension corresponds to the states from which transitions occur, the second dimension corresponds to the states to which transition occurs, and the third dimension gives the event times.

- `cov`: The covariance matrix. Each cell of the matrix gives the estimated covariance between the transition probabilities indicated by the rownames and the colnames, respectively.

- `time`: Times at which the empirical transition matrix is computed

- `s` and `t`: Same as in the function call.

- `trans`: A data frame with two columns `from` and `to` giving the possible transitions.

- `state.names` : Same as in the function call

- `cens.name`: *ditto.*

- `n.risk`: A matrix giving the number of individuals at risk of experiencing a transition.

- `n.event`: An array giving at each times in $(s, t]$ the number of transitions. This is also a 3-dimensional array that is read in the same manner as `est`.

Six methods exist for `etm` objects:

- `print`: Prints details about the multi-state model, e.g., the number of absorbing and transient states, the possible transitions, and the estimates of $\mathbf{P}(s, t)$ and $\mathrm{cov}(\hat{\mathbf{P}}(s, t))$.

- `summary`: The `summary` method returns a list of data frames, one for each transition. Each data frame contains the transition probability estimates, their variance, the number of individuals at risk, the number of events, and the lower and upper bounds of the pointwise confidence intervals.

- `xyplot`: High level trellis function to display the transition probabilities over the course of time. By default, all transition probabilities are drawn, one per panel.

- `plot`: A function that draws all the transition probabilities in a default scatterplot.

- `trprob` and `trcov`: Functions used to extract a specific transition probability and covariance, respectively.

The first data set included in the **etm** package is a subset of the SIR-3 data set (Beyersmann *et al.* 2006). SIR-3 was a prospective cohort study on the effect of nosocomial infections on intensive care patients' survival. The subset present in the package focuses on the effect of ventilation on the length of stay in the intensive care unit. A description and a hazard-based analysis of this data subset is found in Allignol *et al.* (2008). The second data set contains data on pregnancy duration of women exposed to coumarin derivatives (Meister and Schaefer 2008; Allignol *et al.* 2010).

3.2. Illustration: The DLI data set

A random sample of the original DLI data set with added random noise will be used for illustration. It contains information on 304 patients who received allogeneic stem cell transplantation for chronic myeloid leukaemia at the Hammersmith Hospital in London between February 1981 and February 2002. Patients' health status can be described by the nine-state model depicted in Figure 4.

All the patients start the study being alive and in complete remission after the transplant (state 0). Patients can die while in complete remission due to treatment complications (state 1) or relapse (state 2). The patients who have relapsed are offered a salvage treatment of donor lymphocyte infusion (DLI), which is a very effective treatment for restoring complete remis-
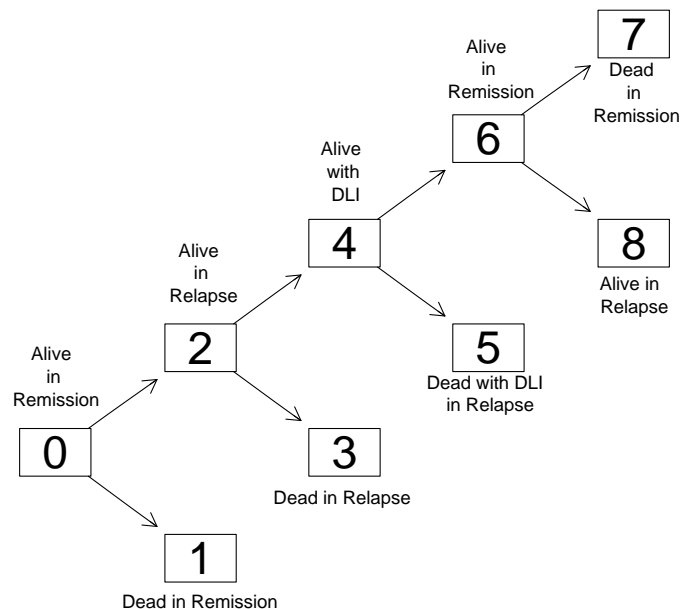

Figure 4: The nine-state model that describes the treatment course of leukaemia patients receiving bone marrow transplant.

sion, though patients may die while waiting for the DLI (state 3). Once they receive the DLI (state 4), patients achieve second complete remission (state 6) or die (state 5). Finally, patients in their second complete remission may then die in remission (state 7) or relapse (state 8).

At the end of the follow-up, 81 patients were alive in remission (state 0). 100 patients died in remission and 123 relapsed. Among them, 18 were censored in state 2 at the end of the study, 40 died while waiting for DLI (state 3), and 65 were offered the salvage therapy. 11 patients were still in state 4 at the end of the follow-up, 10 died before remission (state 5), and 44 achieved complete remission, of whom 38 were still in complete remission at the end of the study, 2 died and 4 relapsed. The patients' course over time may now be studied through transition probabilities. This information is sometimes further condensed into summary quantities, that are functionals of the transition probabilities. A prominent example is the current leukaemia free survival function (Klein *et al.* 2000; Liu *et al.* 2008).

The current leukaemia free survival (CLFS) is defined as the probability that a patient is alive and leukaemia free, either in his first or second post-transplant remission. CLFS is the probability of being in state 0 or 6 at a certain time point t and may be estimated by

$$\widehat{\text{CLFS}}(t) = \hat{P}_{00}(0,t) + \hat{P}_{06}(0,t),$$

where $P_{06}(0,t)$ is the probability that a patient will be alive and in remission after a DLI at time t. The variance of $\widehat{\text{CLFS}}(t)$ may be estimated by

$$\widehat{\text{var}}(\widehat{\text{CLFS}}(t)) = \widehat{\text{var}}(\hat{P}_{00}(0,t)) + \widehat{\text{var}}(\hat{P}_{06}(0,t)) + 2\widehat{\text{cov}}(\hat{P}_{00}(0,t), \hat{P}_{06}(0,t)).$$

In the following, we will demonstrate how the CLFS can be easily computed using the **etm** package, assuming that the model in Figure 4 is time-inhomogeneous Markov.

Below is an excerpt of the data frame with one row per transition,

```
R> load("dli.data.rda")
      id from   to      time distatpr
1      2    0 cens 19.8841843        1
118  256    0    1  9.3126137        1
230  413    0    2  2.3900233        1
231  413    2    3 13.0366784        1
533  614    0    2  0.1792574        2
534  614    2    4  0.5426135        2
535  614    4    6  0.7924446        2
536  614    6    8  4.0103426        2
```

distatpr is the pre-transplant disease status (1: early disease, 2: intermediate or 3: advanced). To use etm, we first need to define the matrix of logical that specifies the possible transitions:

```
R> tra <- matrix(FALSE, 9, 9,
+    dimnames = list(as.character(0:8), as.character(0:8)))
R> tra[1, 2:3] <- TRUE
R> tra[3, 4:5] <- TRUE
R> tra[5, 6:7] <- TRUE
R> tra[7, 8:9] <- TRUE
```

Figure 5: Transition probabilities.
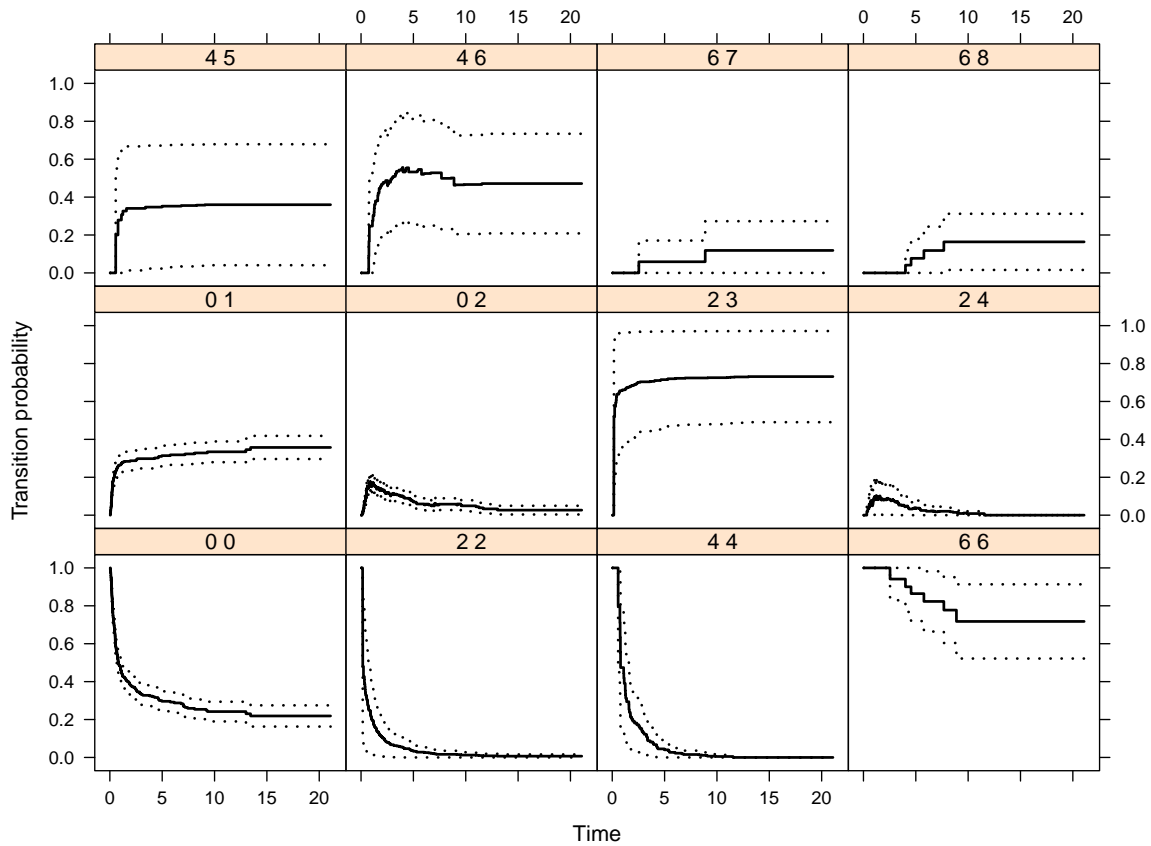
```
        0       1       2       3       4       5       6       7       8
0  FALSE    TRUE    TRUE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE
1  FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE
2  FALSE   FALSE   FALSE    TRUE    TRUE   FALSE   FALSE   FALSE   FALSE
3  FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE
4  FALSE   FALSE   FALSE   FALSE   FALSE    TRUE    TRUE   FALSE   FALSE
5  FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE
6  FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE    TRUE    TRUE
7  FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE
8  FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE
```

Rows of `tra` designate the states from which a transition may occur (e.g., the first row corresponds to state 0, the second row to state 1, and so on), whereas columns represent states to which a transition may occur. For instance, the only possible transitions out of state 0 are towards states 1 or 2. As a consequence, only the second and the third entry of the first row of `tra` are TRUE. Next, we compute the empirical transition matrix with arguments data frame, state names, `tra`, the censoring indicator and `s`. Without any specification, `t` will be the last data time:
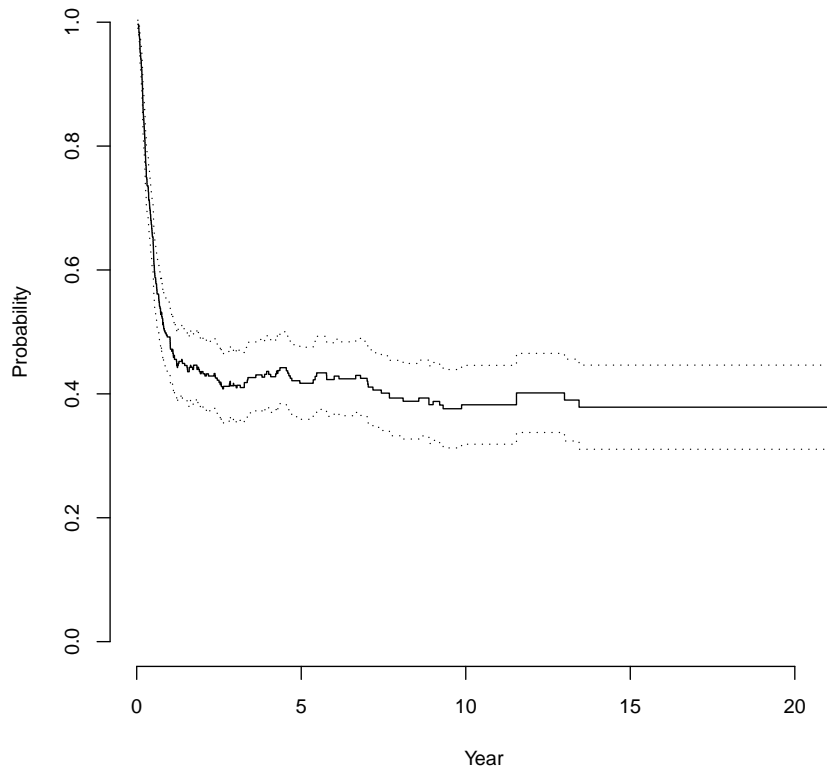
Figure 6: The CLFS curve along with 95% pointwise confidence intervals.

```
R> library("etm")
R> dli.etm <- etm(dli.data, as.character(0:8), tra, "cens", s = 0)
```

We can have a first look at the transition probabilities using the `xyplot` method that by default plots, along with pointwise confidence intervals, the transition probabilities for all possible direct transitions.

```
R> xyplot(dli.etm)
```

The resulting plot is displayed in Figure 5. The bottom left panel displays the leukaemia-free survival function, i.e., the probability to be alive in first remission and leukaemia free. The bottom right panel shows $P_{66}(0, t)$, which provides insights into the effectiveness of the DLI treatment. But the results are further condensed using the CLFS, that is easily obtained once the empirical transition matrix is estimated:

```
R> clfs <- trprob(dli.etm, "0 0") + trprob(dli.etm, "0 6")
R> var.clfs <- trcov(dli.etm, "0 0") + trcov(dli.etm, "0 6") +
+    2 * trcov(dli.etm, c("0 0", "0 6"))
```
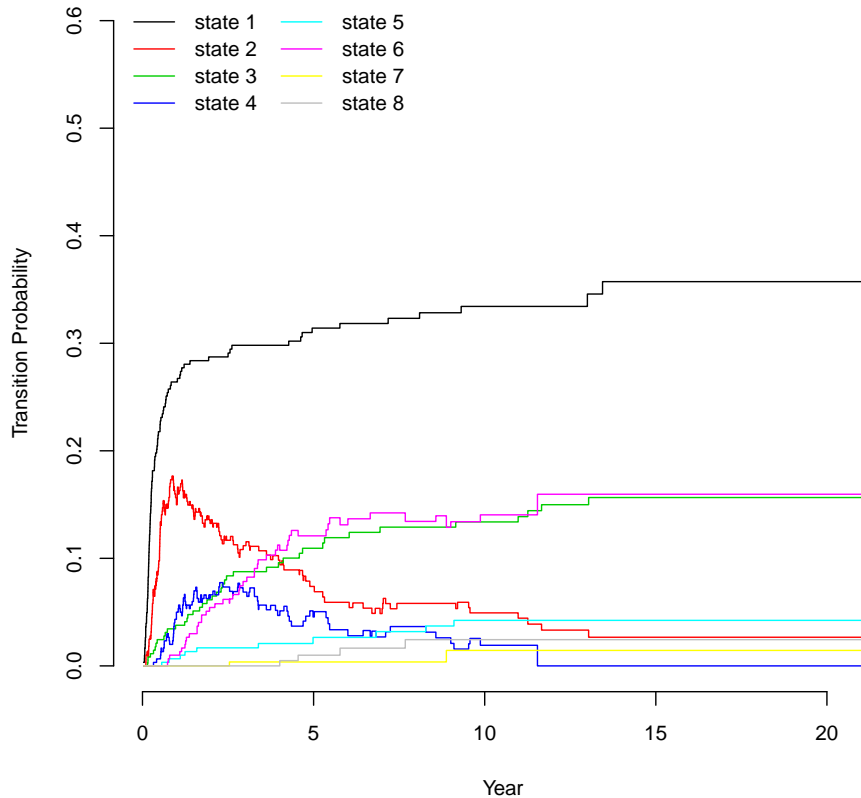
Figure 7: Estimated probabilities of being in states 1 to 8 at time t.

Figure 6 displays the CLFS curve with pointwise confidence intervals. At the end of the follow-up, we observe less than 40% of patients being alive and leukaemia-free (either in first or second remission).

Another useful display that helps to understand how the patients' status develops over time is to plot $\hat{P}_{0j}(0,t)$, $j = 1, \ldots, 8$, *i.e.*, the probabilities of being in each state at a given time point (see Figure 7).

```
R> choice <- rownames(dli.etm$cov)[(grep("^0", rownames(dli.etm$cov)))][-1]
R> plot(dli.etm, tr.choice = choice, col = seq_along(choice), xlab = "Year",
+    lty = rep(1, length(choice)), bty = "n", ylim = c(0, 0.6),
+    curvlab = paste("state", as.character(1:8)), ncol = 2)
```

One may want to compute the CLFS stratified on baseline covariates. The `etm` function is not meant to include covariates, although discrete, potentially time-dependent covariates can be included in the model, making it a joint model, similar as in Andersen (1986). Practically, the `etm` function can be executed using a subset of the data frame. For instance, say we want to compare the CLFS according to the disease stage before transplant, which can be
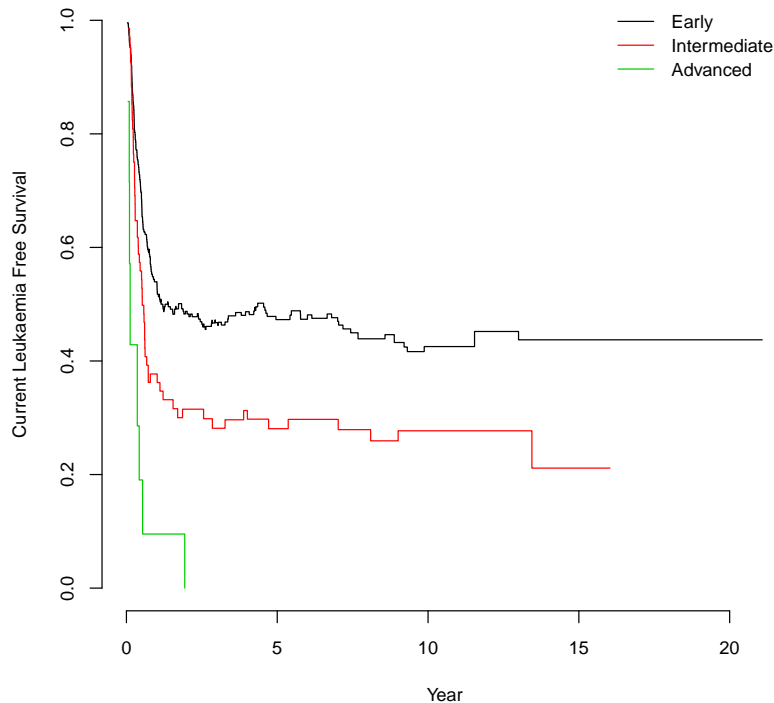
Figure 8: CLFS curves according to the disease stage before transplant.

either early (first chronic phase), intermediate (accelerated phase or greater than or equal to second chronic phase), or advanced (blast phase). The following calls compute the empirical transition matrix stratified on the disease status:

```
R> tr.early <- etm(dli.data[dli.data$distatpr == 1, ], as.character(0:8),
+    tra, "cens", 0, cova = FALSE)
R> tr.int <- etm(dli.data[dli.data$distatpr == 2, ], as.character(0:8),
+    tra, "cens", 0, cova = FALSE)
R> tr.adv <- etm(dli.data[dli.data$distatpr == 3, ], as.character(0:8),
+    tra, "cens", 0, cova = FALSE)
```

Figure 8 displays the CLFS stratified on the disease stage before transplant. As expected, patients in the early stage have a better prognosis, with approximately half of them being in complete remission at the end of the study. In contrast, no patients with an advanced disease were in complete remission at the end of the follow up.

4. Discussion

The **etm** package enables to easily estimate and display the transition probabilities from any time-inhomogeneous Markov multi-state model without covariates. This may be further use-

ful for summary quantities that are functionals of these transition probabilities, as illustrated in the DLI example. As stated above, the covariance estimator that is implemented in **etm** is of the Greenwood type. We just note that estimating the covariance matrix can be computationally expensive. We hope that this package will encourage applied researchers to use multi-state modelling more frequently, as it provides a relevant way to study complex event histories.

We also wish to mention that after **etm** was released on CRAN, the **mstate** package (de Wreede *et al.* 2011) has also been made available on CRAN. Among other things, the package also allows to compute the empirical transition matrix with covariances, both with and without covariates.

Finally, we note that the empirical transition matrix remains a consistent estimator of the state occupation probabilities $P_{ii}(s, t)$, $i \in \mathcal{S}$, in a non-Markov model if censoring is entirely random (Datta and Satten 2001). However, variance estimation should probably be based on resampling in such a situation (Glidden 2002).

Acknowledgments

References

Aalen OO (1978). "Nonparametric Estimation of Partial Transition Probabilities in Multiple Decrement Models." *The Annals of Statistics*, **6**, 534–545.

Aalen OO, Johansen S (1978). "An Empirical Transition Matrix for Non-Homogeneous Markov Chains Based on Censored Observations." *Scandinavian Journal of Statistics*, **5**, 141–150.

Allignol A, Beyersmann J, Schumacher M (2008). "**mvna**: An R Package for the Nelson-Aalen Estimator in Multistate Models." *R News*, **8**(2), 48–50. URL http://CRAN.R-project.org/doc/Rnews/.

Allignol A, Schumacher M, Beyersmann J (2010). "A Note On Variance Estimation of the Aalen-Johansen Estimator of the Cumulative Incidence Function in Competing Risks, with a View towards Left-Truncated Data." *Biometrical Journal*, **52**, 126–137.

Andersen P, Keiding N (2002). "Multi-State Models for Event History Analysis." *Statistical Methods in Medical Research*, **11**(2), 91–115.

Andersen P, Pohar Perme M (2008). "Inference for Outcome Probabilities in Multi-State Models." *Lifetime Data Analysis*, **14**(4), 405–431.

Andersen PK (1986). "Time-Dependent Covariates and Markov Processes." In SH Moolgavkar, RL Prentice (eds.), *Modern Statistical Methods in Chronic Disease Epidemiology*, pp. 82–103. John Wiley & Sons.

Andersen PK, Borgan Ø, Gill RD, Keiding N (1993). *Statistical Models Based on Counting Processes.* Springer-Verlag, New-York.

Beyersmann J, Gastmeier P, Grundmann H, Baerwolff S, Geffers C, Behnke M, Rueden H, Schumacher M (2006). "Use of Multistate Models to Assess Prolongation of Intensive Care Unit Stay Due to Nosocomial Infection." *Infection Control and Hospital Epidemiology*, **27**(5), 493–499.

Datta S, Satten G (2001). "Validity of the Aalen–Johansen Estimators of Stage Occupation Probabilities and Nelson–Aalen Estimators of Integrated Transition Hazards for Non-Markov Models." *Statistics and Probability Letters*, **55**(4), 403–411.

de Wreede LC, Fiocco M, Putter H (2011). "**mstate**: An R Package for the Analysis of Competing Risks and Multi-State Models." *Journal of Statistical Software*, **38**(7), 1–30. URL http://www.jstatsoft.org/v38/i07/.

Fleming T (1978a). "Asymptotic Distribution Results in Competing Risks Estimation." *The Annals of Statistics*, **6**(5), 1071–1079.

Fleming T (1978b). "Nonparametric Estimation for Nonhomogeneous Markov Processes in the Problem of Competing Risks." *The Annals of Statistics*, **6**, 1057–1070.

Gill R, Johansen S (1990). "A Survey of Product-Integration with a View toward Application in Survival Analysis." *The Annals of Statistics*, **18**(4), 1501–1555.

Glidden D (2002). "Robust Inference for Event Probabilities with Non-Markov Event Data." *Biometrics*, **58**(2), 361–368.

Gray RJ (2010). ***cmprsk**: Subdistribution Analysis of Competing Risks.* R package version 2.2-1, URL http://CRAN.R-project.org/package=cmprsk.

Hougaard P (2000). *Analysis of Multivariate Survival Data.* Springer-Verlag, New York.

Klein J, Shu Y (2002). "Multi-State Models for Bone Marrow Transplantation Studies." *Statistical Methods in Medical Research*, **11**(2), 117–139.

Klein J, Szydlo R, Craddock C, Goldman J (2000). "Estimation of Current Leukaemia-Free Survival Following Donor Lymphocyte Infusion Therapy for Patients with Leukaemia who Relapse after Allografting: Application of a Multistate Model." *Statistics in Medicine*, **19**, 3005–3016.

Liu L, Logan B, Klein J (2008). "Inference for Current Leukemia Free Survival." *Lifetime Data Analysis*, **14**(4), 1–15.

Meister R, Schaefer C (2008). "Statistical Methods for Estimating the Probability of Spontaneous Abortion in Observational Studies–Analyzing Pregnancies Exposed to Coumarin Derivatives." *Reproductive Toxicology*, **26**(1), 31–35.

Moon H, Lee JJ, Ahn H, Nikolova RG (2002). "A Web-Based Simulator for Sample Size and Power Estimation in Animal Carcinogenicity Studies." *Journal of Statistical Software*, **7**(13), 1–36. URL http://www.jstatsoft.org/v07/i13.

Putter H, Fiocco M, Geskus R (2007). "Tutorial in Biostatistics: Competing Risks and Multi-State Models." *Statistics in Medicine*, **26**(11), 2389–2430.

R Development Core Team (2010). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/.

Therneau T, Lumley T (2010). **survival**: *Survival Analysis Including Penalised Likelihood*. R package version 2.36-1, URL http://CRAN.R-project.org/package=survival.

Wangler M, Beyersmann J, Schumacher M (2006). "**changeLOS**: An R-Package for Change in Length of Hospital Stay Based on the Aalen-Johansen Estimator." *R News*, **6**(2), 31–35. URL http://CRAN.R-project.org/doc/Rnews/.

Affiliation:

Arthur Allignol, Jan Beyersmann
Freiburg Center for Data Analysis and Modeling
University of Freiburg
Institute for Medical Biometry and Informatics
University Medical Centre, Freiburg
79104 Freiburg, Germany
E-mail: arthur.allignol@fdm.uni-freiburg.de,
        Jan.Beyersmann@fdm.uni-freiburg.de

Martin Schumacher
Institute for Medical Biometry and Informatics
University Medical Centre, Freiburg
79104 Freiburg, Germany
E-mail: ms@imbi.uni-freiburg.de