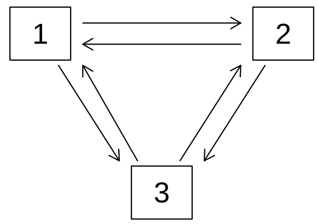
Generating Multistate Data with Intermittent Observation

Here we consider a reversible 3-state model with state-space diagram below.



We consider a time-homogeneous process with transition intensity matrix $Q(t) = Q$ of the form

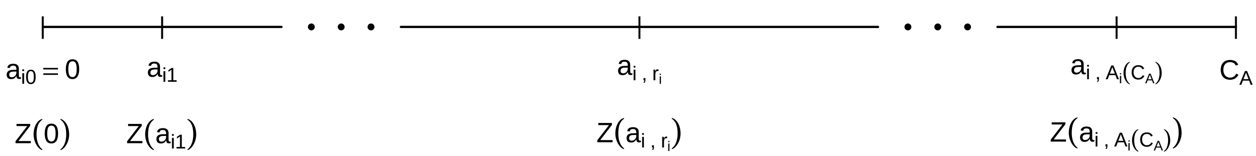
$$Q = \begin{bmatrix} -(\lambda_{12} + \lambda_{13}) & \lambda_{12} & \lambda_{13} \\ \lambda_{21} & -(\lambda_{21} + \lambda_{23}) & \lambda_{23} \\ \lambda_{31} & \lambda_{32} & -(\lambda_{31} + \lambda_{32}) \end{bmatrix}.$$

To generate data when such a process is under intermittent observation over $(0, C_A]$ we proceed as follows.

1. We generate the assessment times according to a time homogeneous Poisson process with rate ρ; Let A_{ir} denote the random (potentially latent) time of the rth assessment for individual i. The waiting times $\Delta A_{ir} = A_{ir} - A_{i,r-1}$, $r = 1, \ldots,$ are independent and identically exponentially distributed and so given $a_{i0} = 0$ we generate $a_{i1}, a_{i2}, \ldots$ as follows. For $r = 1$ we generate

   $$\Delta A_{ir} \sim \text{EXP}(\text{rate} = \rho)\,.$$

   If Δa_{ir} is the realization, given $a_{i,r-1}$ we set $a_{ir} = \Delta a_{ir} + a_{i,r-1}$. If $a_{ir} > C_A$, then we stop and assessment $r-1$ is the last occuring at time $a_{i,r-1}$. Otherwise increment $r = r + 1$ and repeat this step. If we let $\{A_i(s), s > 0\}$ denote the counting process for the visits for individual i then $A_i(C_A)$ is the number of assessments over $(0, C_A]$.



2. At $a_{i0} = 0$, we generate $Z_i(0)$ based on an initial distribution where

$$Z_i(0) \sim \text{Multi}(1,\ (\pi_1, \pi_2, \pi_3))$$

where $\pi_k = P(Z_i(0) = k)$ for $k = 1, 2, 3$ and $\pi_1 + \pi_2 + \pi_3 = 1$.

3. Given $(Z_i(a_{i,r-1}), a_{i,r-1})$ and a_{ir} we generate $Z_i(a_{ir})$ as multinomial with

$$Z_i(a_{ir})|Z_i(a_{i,r-1}), a_{i,r-1}, a_{ir} \sim \text{Multi}(1,\ (P_{Z_i(a_{i,r-1}),1}(a_{i,r-1}, a_{ir}),\ P_{Z_i(a_{i,r-1}),2}(a_{i,r-1}, a_{ir}),\ P_{Z_i(a_{i,r-1}),3}(a_{i,r-1}, a_{ir})))$$

for $r = 1, 2, \ldots, A_i(C_A)$,

4. We repeat these steps for each individual i, $i = 1, \ldots, m$.

In the accompanying R code, the input parameters for the functions are

- $m = 1000$
- $C_A = 1$
- $(\pi_1, \pi_2, \pi_3) = (0.4,\ 0.3,\ 0.3)$
- $\lambda_{12} = 0.6$, $\lambda_{13} = 0.4$, $\lambda_{21} = 0.3$, $\lambda_{23} = 0.2$, $\lambda_{31} = 0.2$ and $\lambda_{32} = 0.1$
- $\rho = 5$

The output dataframe consists of columns `id`, `times` and `states` where

$$\texttt{times} = (a_{i0} = 0,\ a_{i1},\ \ldots,\ a_{i,A_i(C_A)},\ C_A)$$

and

$$\texttt{states} = (Z_i(0),\ Z_i(a_{i1}),\ \ldots,\ Z_i(a_{i,A_i(C_A)}),\ 999)\,.$$

Let 999 indicate a censoring state.