Simulating longitudinal data for time-to-event analysis in continuous time

We simulate a cohort of patients who initiate treatment at time $t = 0$, denoted by $A(0) = 1$ and who are initially stroke-free, $L(0) = 0$. All individuals are followed for up to $\tau_{\text{end}} = 730$ days or until death. Initially, we do not consider censoring or competing events. During follow-up, patients may experience (at most) one stroke, stop treatment (irreversibly), and die, that is $N^x(t) \leq 1$ for $x = a, \ell, y$. With these assumptions $K = 2$ in the main note (at most two non-terminal events). The primary outcome is the *risk of death within 2 years.*

As in the note, we assume that a treatment event and a covariate event cannot happen at the same time, which is not a significant limitation. Our observations consist of

$$O = \left(T_{(3)}, \Delta_{(3)}, A\left(T_{(2)}\right), L\left(T_{(2)}\right), T_{(2)}, \Delta_{(2)}, A\left(T_{(1)}\right), L\left(T_{(1)}\right), T_{(1)}, \Delta_{(1)}, A(0), L(0), \text{age}\right),$$

where $T_{(k)}$ is the time of the k'th event, $\Delta_{(k)} \in \{\ell, a, y, c\}$ (stroke, visit, death, censored), $A\left(T_{(k)}\right)$ is the treatment status at time $T_{(k)}$, and $L\left(T_{(k)}\right)$ is the value of the covariate at time $T_{(k)}$. We reserve c for administrative censoring, corresponding to the event happening after the end of the study period, τ_{end}. Note that we let $T_{(k)} = \infty$ if the k'th event cannot happen (because the previous event was a terminal event or the end of the study period was reached).

Then, we generate the baseline variables as follows

$$\begin{aligned} &\text{age} \sim \text{Unif}(40, 90), \\ &L = 0, \\ &A(0) = 1, \end{aligned}$$

Now we describe the simulation mechanism corresponding to the first event that can happen. To allow for administrative censoring, let $T^c_{(1)} = 2$. We define $T^a_{(1)}$ such that the patient can be expected to go to the doctor within the first year, if the two other events have not occurred first. Let $\text{Exp}(\lambda)$ denote the exponential distribution with rate $\lambda \geq 0$. We let $\lambda = 0$ correspond to the case that the event cannot happen, i.e., $T^x_{(1)} = \infty$. As the first event, we draw

$$\begin{aligned} &T^x_{(1)} \sim \text{Exp}\left(\lambda^x_0 \exp\left(\beta^x_{0,\text{age}} \text{ age}\right)\right), \qquad x = \ell, y \\ &T^a_{(1)} \sim 1 + \mathcal{N}(0, \delta) \\ &\Delta_{(1)} = \text{argmin}_{x=a,\ell,y,c} T^x_{(1)} \\ &T_{(1)} = T^{\Delta_{(1)}}_{(1)} \\ &A\left(T_{(1)}\right) \mid T_{(1)} = t, \text{age} = x \begin{cases} \sim \text{Bernoulli}(\text{expit}\left(\alpha_{00} + \alpha_{0,\text{ age}} x\right) \text{ if } \Delta_{(1)} = a \\ 1 \text{ otherwise} \end{cases} \\ &L\left(T_{(1)}\right) = 1, \end{aligned}$$

Note that we simulate from a “competing event setup” by defining latent variables $T^x_{(1)}$ for each possible event type x.

We now describe the second event that can happen. If the first event was a terminal event – either outcome or administrative censoring – we stop and do not generate more data for this patient. Now, we let $S_{(2)}$ denote the time between $T_{(1)}$ and the second event $T_{(2)}$, i.e., $S_{(2)} = T_{(2)} - T_{(1)}$. As we required that $N^x(t) \leq 1$, if the first event was a stroke, we cannot have a second stroke, and if the first event was a visit, we cannot have a second visit. If the first event was a stroke, the doctor visit is likely to happen soon after, so we simulate the corresponding latent time as an exponential random variable. We will then generate the second event time $T_{(2)}$ as follows:

$$\begin{aligned}
&S_{(2)}^{\ell} \sim \operatorname{Exp}\left(\lambda_1^{\ell} \exp\left(\beta_{1,\text{age}}^{\ell} \text{ age} + \beta_{1,A}^{\ell}\left(1 - A\left(T_{(1)}\right)\right)\right)\mathbb{1}\left\{\Delta_{(1)} = a\right\}\right) \\
&S_{(2)}^{y} \sim \operatorname{Exp}\left(\lambda_1^{y} \exp\left(\beta_{1,\text{age}}^{y} \text{ age} + \beta_{1,A}^{y}\left(1 - A\left(T_{(1)}\right)\right) + \beta_{1,L}^{\ell} L\left(T_{(1)}\right)\right)\right) \\
&S_{(2)}^{c} = 2 - T_{(1)} \\
&S_{(2)}^{a} \sim \operatorname{Exp}\left(\gamma_0 \exp\left(\gamma_{\text{age}} \text{ age}\right)\mathbb{1}\left\{\Delta_{(1)} = \ell\right\}\right) \\
&\Delta_{(2)} = \operatorname{argmin}_{x=a,\ell,y,c} S_{(2)}^{x} \\
&T_{(2)} = T_{(1)} + S_{(2)}^{\Delta_{(2)}} \\
&A\left(T_{(2)}\right) \mid T_{(2)} = t, \text{age} = x, A\left(T_{(1)}\right) = a_1, L\left(T_{(1)}\right) = l_1 \\
&\qquad = \begin{cases} \sim \text{Bernoulli}(\operatorname{expit}\left(\alpha_{10} + \alpha_{1,\text{ age}} x\right) \text{ if } \Delta_{(2)} = a \\ 1 \text{ otherwise} \end{cases} \\
&L\left(T_{(2)}\right) = 1.
\end{aligned}$$

In the following, we assumed that the previous event times have no influence on anything, only the "marks". However, this may be unrealistic, as the effect of a stroke on mortality may naturally decrease over time.

Finally, we let $T_{(3)} = S_{(3)} + T_{(2)}$ denote the time of the third event, if it can happen. We define the time $S_{(3)}$ as follows:

$$\begin{gathered}
S_{(3)}^{y} \sim \operatorname{Exp}\left(\lambda_2^{y} \exp\left(\beta_{2,\text{age}}^{y} \text{ age} + \beta_{2,A}^{y}\left(1 - A\left(T_{(2)}\right)\right) + \beta_{2,L}^{y} L\left(T_{(2)}\right)\right)\right) \\
S_{(3)}^{c} = 2 - T_{(2)} \\
\Delta_{(3)} = \operatorname{argmin}_{x=y,c} S_{(3)}^{x} \\
T_{(3)} = T_{(2)} + S_{(3)}^{\Delta_{(3)}}.
\end{gathered}$$

Here, we furthermore make the assumption that it does not matter whether the patient had a stroke first and then visited the doctor, or visited the doctor first and then had a stroke.

When the static intervention is applied, we put $A\left(T_{(k)}\right) = 1$ for each $k = 1, ..., K$. It is not too difficult to see that the likelihood factorizes as in Rytgaard et al. (2022) corresponding to the intervention that we are interested in (see e.g., Theorem II.7.1 of Andersen et al. (1993)).

Plain Language Summary (for Clinical Audience)

We simulate patients who all begin treatment and are initially healthy. Over two years, they may have a stroke, stop treatment (only at doctor visits), or die. A routine doctor visit is scheduled about a year after treatment begins, unless a stroke happens first, in which case a visit is likely to occur soon after. Doctors are less likely to stop treatment after a stroke. The chance of dying depends on whether the patient has had a stroke and whether they are still on treatment.

Bibliography

Andersen, P. K., Borgan, Ø., Gill, R. D., & Keiding, N. (1993). *Statistical Models Based on Counting Processes*. Springer US. https://doi.org/10.1007/978-1-4612-4348-9

Rytgaard, H. C., Gerds, T. A., & Laan, M. J. van der. (2022). Continuous-Time Targeted Minimum Loss-Based Estimation of Intervention-Specific Mean Outcomes. *The Annals of Statistics*, *50*(5), 2469–2491. https://doi.org/10.1214/21-AOS2114