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

Each observation $O = \left(T_{(K)}, \Delta_{(K)}, A\left(T_{(K-1)}\right), L\left(T_{(K-1)}\right), T_{(K-1)}, \Delta_{(K-1)}, ..., A(0), L(0)\right)$ is generated in the following way. Recall from the main note that we put $\mathcal{F}_{T_{(k)}} = \sigma\left(T_{(k)}, \Delta_{(k)}, A\left(T_{(k-1)}\right), L\left(T_{(k-1)}\right)\right) \vee \mathcal{F}_{T_{(k-1)}}$.

Let $\pi_k\left(t, \mathcal{F}_{T_{(k-1)}}\right)$ be the probability of being treated at the k'th event given $\Delta_{(k)} = a, T_{(k)} = t$, and $\mathcal{F}_{T_{(k-1)}}$. Similarly, let $\mu_k\left(t, \cdot, \mathcal{F}_{T_{(k-1)}}\right)$ be the probability measure for the covariate value given $\Delta_{(k)} = \ell, T_{(k)} = t$, and $\mathcal{F}_{T_{(k-1)}}$. Let also $\Lambda_k^x\left(\mathrm{d}t, \mathcal{F}_{T_{(k-1)}}\right)$ be the cumulative cause-specific hazard measure for the k'th event and cause x given $\mathcal{F}_{T_{(k-1)}}$, where $x = a, \ell, d, y, c$. At baseline, we let $\pi_0(L(0))$ be the probability of being treated given $L(0)$ and $\mu_0(\cdot)$ be the probability measure for the covariate value.

We let $L(t)$ consist of the covariates *age*, *sex*, $L_1(t)$, $L_2(t)$ (e.g., recurrent events). Then we generate the baseline variables as follows

$$
\begin{aligned}
&\text{age} \sim \text{Unif}(40, 90) \\
&\text{sex} \sim \text{Bernoulli}(0.4) \\
&L_1(0) \sim \text{Bernoulli}(0.4) \\
&L_2(0) \sim \text{Bernoulli}(0.25) \\
&A(0) \sim \text{Bernoulli}\left(\text{expit}\left(\left(\beta_0^a\right)^T \mathcal{F}_0^{\text{-A}} + \beta_0^{a,*}\right)\right),
\end{aligned}
$$

where $\mathcal{F}_0^{\text{-A}} = (\text{age}, \text{sex}, L_1(0), L_2(0))$.

Then, the observation is drawn iteratively as follows,

$$
\begin{aligned}
&S_{(k)}^x \mid \mathcal{F}_{T_{(k-1)}} = f_{t_{k-1}} \sim \text{Exp}\left(\lambda_k^x \exp\left(\left(\beta_k^x\right)^T f_{t_{k-1}}\right)\right), x = a, \ell, d, y, c \\
&\Delta_{(k)} = x \text{ if } S_{(k)}^x < S_{(k)}^z \text{ for all } z \neq x \\
&T_{(k)} = T_{(k-1)} + S_{(k)}^x \text{ if } \Delta_{(k)} = x \\
&L^* \mid T_{(k)}, \mathcal{F}_{T_{(k-1)}} = f_{t_{k-1}} \sim \text{Bernoulli}(\text{expit}(\alpha_k^L)^T f_{t_{k-1}} + \alpha_k^{L,*}) \\
&L_1(0) = \begin{cases} L_1(k-1) + L^* \text{ if } \Delta_{(k)} = \ell \text{ and } k < K \\ L_1(k-1) \text{ otherwise} \end{cases} \\
&L_2(0) = \begin{cases} L_2(k-1) + L^* \text{ if } \Delta_{(k)} = \ell \text{ and } k < K \\ L_2(k-1) \text{ otherwise} \end{cases} \\
&A\left(T_{(k)}\right) = \text{Bernoulli}(\text{expit}(\left(\alpha^A k\right)^T f_{t_{k-1}} + \alpha_k^{A,*})) \text{ if } \Delta_{(k)} = a
\end{aligned}
$$

where $\text{Exp}(\lambda)$ denotes the exponential distribution with rate λ. When the static intervention is applied, we put $A\left(T_{(k)}\right) = 1$ for each $k = 1, ..., K$. When the uncensored data argument is used, we put $S_{(k)}^c = \infty$. So the parameters we can vary are the α's, β's, and λ's. A limitation of the current implementation is that the Markov assumption is used for the time-varying variables, i.e., $S_{(k)}^x$ depends only on $\mathcal{F}_{T_{(k-1)}}$ through $\left(A\left(T_{(k-1)}\right), L\left(T_{(k-1)}\right), T_{(k-1)}, \Delta_{(k-1)}\right)$.