A new iterated conditional expectations estimator for longitudinal causal effects in continuous time

Johan Sebastian Ohlendorff[1], Anders Munch[1], and Thomas Alexander Gerds[1],
[1] Section of Biostatistics, University of Copenhagen

UNIVERSITY OF COPENHAGEN

Motivation

- In medical research, the estimation of causal effects of treatments over time is often of interest.
- Continuous-time inference allows for data that is more closely aligned with the data collection process (Table 1). Moreover, discrete time approaches usually require the discretization of time, leading to a loss of information.
- There is a scarcity of (applied) literature on the estimation of longitudinal causal effects in continuous time. Rytgaard et al. (2022) considered a targeted minimum-loss based estimator based on iterated conditional expectations (Figure 1) for estimating causal effects flexibly. Recently, Ryalen (2024) proposed a general identification result for longitudinal causal effects in continuous time. We extend upon these works and provide a new feasible iterated conditional expectations estimator (Figure 2) for the estimation of longitudinal causal effects in continuous time.

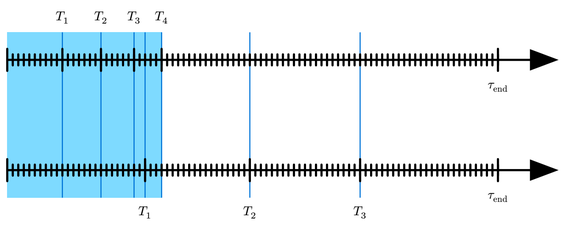

Figure 1: The figure illustrates the sequential regression approach given in Rytgaard et al. (2022) for two observations.

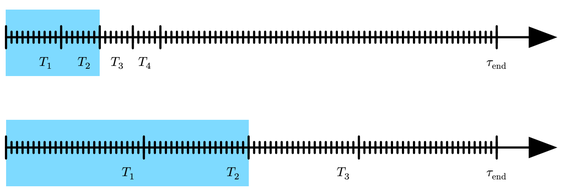

Figure 2: The figure illustrates the sequential regression approach proposed in this article.

id	time	event
1	3	side effect
1	8	primary event
2	10	primary event
3	2	side effect
3	5	treatment shift
3	7	censoring
	...	

Table 1: An example of a longitudinal dataset from electronic health records or a clinical trial. Events are registered at irregular/subject-specific time points.

Setting

Let $\left(N^a(t), A(t), N^\ell(t), L(t), N^y(t), N^d(t), N^c(t)\right)$[1] be a stochastic (jump) processes observed in $[0, \tau_{\text{end}}]$, consisting of a counting process for treatment visits, treatment values, a counting process for treatment covariate measurements, covariate values, and counting processes for the primary event, competing event, and censoring, respectively. Furthermore, $A(t) \in \{0, 1\}$ and $L(t) \in \mathcal{L}$, where $\mathcal{L} \subseteq \mathbb{R}^d$ is a finite set.

Assumption 1: In the time interval $[0, \tau_{\text{end}}]$ there are at most $K - 1 < \infty$ many changes of treatment and covariates in total for a single individual.

Assumption 2: The counting processes N^a, N^ℓ, N^y, N^d, and N^c have with probability 1 no jump times in common.

Under these assumptions, the observed data can be written in the form

$$O = \mathcal{F}_{T_{(K)}}$$

where

$$\mathcal{F}_{T_{(k)}} = \left(T_{(k)}, \Delta_k, A\left(T_{(k)}\right), L\left(T_{(k)}\right)\right) \vee \mathcal{F}_{T_{(k-1)}} \text{ and } \mathcal{F}_0 = (L_0, A_0)$$

Assumption 3: For each $k \in \{1, ..., K\}$, $T_{(k)} \mid \mathcal{F}_{T_{(k-1)}} \ll m$[2], $A\left(T_{(k)}\right) \mid T_{(k)} = t, \Delta_k = a, \mathcal{F}_{T_{(k-1)}} \ll \nu_a$, and $L\left(T_{(k)}\right) \mid T_{(k)} = t, \Delta_k = \ell, \mathcal{F}_{T_{(k-1)}} \ll \nu_\ell$.

Target parameter

Let $\tilde{T}_k^{\mathbf{1}}$ and $\tilde{\Delta}_k^{\mathbf{1}}$ be the counterfactual event time and indicator for the k'th had the patient stayed on treatment and initially received treatment (and not been censored). Our target parameter $\Psi_\tau^g : \mathcal{M} \to \mathbb{R}$ is the mean interventional absolute risk at time τ given the intervention plan g.

$$\Psi_\tau^g(P) = \mathbb{E}_P\left[\sum_{k=1}^{K} \mathbb{1}\left\{\tilde{T}_k^{\mathbf{1}} \leq \tau, \tilde{\Delta}_k^{\mathbf{1}} = y\right\}\right].$$

Identification

We consider identification conditions in Theorem 3 of Ryalen (2024). These are stated in our present uncensored setting. Let $\tilde{Y}_t = \left(\mathbb{1}\left\{\tilde{T}_1^{\mathbf{1}} \leq t, \tilde{\Delta}_1^{\mathbf{1}} = y\right\}, ..., \mathbb{1}\left\{\tilde{T}_K^{\mathbf{1}} \leq t, \tilde{\Delta}_K^{\mathbf{1}} = y\right\}\right)$ and $T^a = \inf\{t > 0 : A(t) \neq 1\}$. For each $k \in \{1, ..., K\}$, we need:

- **Consistency**:

$$\mathbb{1}\left\{\tilde{T}_k^{\mathbf{1}} \leq t, \tilde{\Delta}_k^{\mathbf{1}} = y\right\}\mathbb{1}\{T^a > t, A(0) = 1\} = \mathbb{1}\{T_k \leq t, \Delta_k = y\}\mathbb{1}\{T^a > t, A(0) = 1\}$$

  for $t \in [0, \tau_{\text{end}}]$.
- **Exchangeability**:

$$A\left(T_{(k)}\right) \perp \left(\left(\mathbb{1}\left\{\tilde{T}_{k+1}^{\mathbf{1}} \leq t, \tilde{D}_{k+1}^{\mathbf{1}} = y\right\}, ..., \mathbb{1}\left\{\tilde{T}_K^{\mathbf{1}} \leq t, \tilde{D}_K^{\mathbf{1}} = y\right\}\right)\right)_{t \in [0, \tau_{\text{end}}]} \mid \Delta_k = a, \mathcal{F}_{T_{(k-1)}}$$

  (and

$$\lambda^a\left(t \mid \mathcal{F}_{T_{(k-1)}} \vee \left(\tilde{Y}_t\right)_{t \in [0, \tau_{\text{end}}]}\right)$$

$$= \lim_{h \to 0} \frac{P\left(t \leq T_{(k)} < t + h, \Delta_k = a \mid T_{(k)} \geq t, \mathcal{F}_{T_{(k-1)}}, \left(\tilde{Y}_t\right)_{t \in [0, \tau_{\text{end}}]}\right)}{h}$$

  does not depend on $\left(\tilde{Y}_t\right)_{t \in [0, \tau_{\text{end}}]}$).
- **Positivity**: The weights

$$w_k(f_{k-1}, t_k) = \frac{\mathbb{1}\{a_0 = 1\}}{\pi_0(l_0)} \prod_{j=1}^{k-1} \left(\frac{\mathbb{1}\{a_j = 1\}}{\pi_j(f_{j-1})}\right)^{\mathbb{1}\{\delta_j = a\}} \mathbb{1}\{t_1 < ... < t_k\}$$

  fulfill $\mathbb{E}_P\left[w_k\left(\mathcal{F}_{T_{(k-1)}}, T_{(k)}\right)\right] = 1$. Here $\pi_0(l) = P(A(0) = 1 \mid L(0) = l_0)$ and $\pi_j(f) = P\left(A\left(T_{(j)}\right) = 1 \mid \Delta_j = a, \mathcal{F}_{T_{(j-1)}} = f\right)$.

Identification formula

Under the assumptions of consistency, exchangeability, and positivity, the target parameter is identified via

$$\Psi_\tau^g(P) = \mathbb{E}_P\left[\sum_{k=1}^{K} w_k\left(\mathcal{F}_{T_{(k-1)}}, T_{(k)}\right)\mathbb{1}\{T_k \leq \tau, \Delta_k = y\}\right].$$

Iterated conditional expectation estimator

The form of the efficient influence function (Bickel et al. (1993)) in this setting suggests the use of a iterated conditional expectations estimator. Let $S^c\left(t \mid \mathcal{F}_{T_{(k)}}\right)$ be the conditional survival function of the censoring time given the k previous events and $\mathcal{F}_{T_{(k)}}^{-A}$ denote the history without the treatment process.

Proposed continuous-time ICE algorithm

- For each event point $k = K, K-1, ..., 1$ (starting with $k = K$):
  1. Obtain $\hat{S}^c\left(t \mid \mathcal{F}_{T_{(k)}}\right)$ by fitting a cause-specific hazard model for the censoring via the interevent time $S_{(k)} = T_{(k)} - T_{(k-1)}$, regressing on $\mathcal{F}_{T_{(k-1)}}$ (among the people who are still at risk after $k-1$ events).
  2. Define the subject-specific weight:

$$\hat{\eta}_k = \frac{\mathbb{1}\left\{T_{(k)} \leq \tau, \Delta_k \in \{a, \ell\}, k < K\right\}\hat{\nu}_k\left(\mathcal{F}_{T_{(k)}}^{-A}, \mathbf{1}\right)}{\hat{S}^c\left(T_{(k)} \mid \mathcal{F}_{T_{(k-1)}}^{-A}, \mathbf{1}\right)}$$

     Then calculate the subject-specific pseudo-outcome

$$\hat{R}_k = \frac{\mathbb{1}\left\{T_{(k)} \leq \tau, \Delta_k = y\right\}}{\hat{S}^c\left(T_{(k)} \mid \mathcal{F}_{T_{(k-1)}}^{-A}, \mathbf{1}\right)} + \hat{\eta}_k$$

     Regress $\hat{R}_k$ on $\mathcal{F}_{T_{(k-1)}}$ on the data with $T_{(k-1)} < \tau$ and $\Delta_k \in \{a, \ell\}$ to obtain a prediction function $\hat{\nu}_{k-1} : \mathcal{H}_{k-1} \to \mathbb{R}_+$.
     - At baseline, we obtain the estimate $\hat{\Psi}_n = \frac{1}{n}\sum_{i=1}^{n} \hat{\nu}_0(L_i(0), 1)$.

Future directions/challenges

- Implementation of the method and application on real data.
- Few individuals may have a high number of events, leading to potentially very small sample sizes in the iterated regressions.

References

Bickel, P. J., Klaassen, C. A., Bickel, P. J., Ritov, Y., Klaassen, J., Wellner, J. A., & Ritov, Y. (1993). *Efficient and adaptive estimation for semiparametric models* (Vol. 4). Johns Hopkins University Press Baltimore.

Ryalen, P. (2024). *On the role of martingales in continuous-time causal inference.*

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

[1] We associate to this process its natural filtration $\mathcal{F}_t$ implicitly defined on a probability space $(\Omega, \mathcal{F}, P)$.
[2] m is the Lebesgue measure on $\mathbb{R}_+$. Also $\ll$ should be understood in terms of the corresponding Markov kernel.