Potential outcomes

Potential outcomes (martingale approach)

Let us consider the setting of Ryalen (2024) and their regularity conditions. Specifically, we will work with an intervention that specifies the treatment decisions but not the timing of treatment visits. We work with Example 4 of Ryalen (2024), in which

$$\pi^*(\varphi, t, \mathrm{d}x) = \delta_{a_0}(\mathrm{d}x),$$

i.e., treatment is always assigned to a_0. To simplify, we work without right-censoring and no covariates. This means that (N^y, N^a), where N^y denotes the counting process on $[0, T]$ for death and N^a random measure for treatment on $[0, T] \times \{a_0, a_1\}$. For this treatment regime, we see that

$$\tau^A = \inf\{t \geq 0 \mid N^a((0, t] \times \{a_1\}) > 0\}.$$

We can associate each of the random measures N^y and N^a with the random measure

$$N(\mathrm{d}(t, m, a)) = N^y(\mathrm{d}t)\delta_y(\mathrm{d}m) + \delta_a(\mathrm{d}m)\Big\{N^a(\mathrm{d}(t) \times \{a_0\})\delta_{a_0}(\mathrm{d}m) + N^a(\mathrm{d}(t) \times \{a_1\})\delta_{a_1}(\mathrm{d}m)\Big\}.$$

This gives rise to a counting process filtration $(\mathcal{F}_t)_{t\geq 0}$ generated by N. We can then find that N has the compensator

$$\Lambda(\mathrm{d}(t, m, a)) = \Lambda^y(\mathrm{d}t)\delta_y(\mathrm{d}m) + \delta_a(\mathrm{d}m)\Big\{\pi_t(\mathcal{F}_{t-})\delta_{a_0}(\mathrm{d}m) + (1 - \pi_t(\mathcal{F}_{t-}))\delta_{a_1}(\mathrm{d}m)\Big\}\Lambda^a(\mathrm{d}(t)),$$

where we can choose π_t to be $\mathcal{F}_t$-predictable. We are interested in the counterfactual mean outcome $\mathbb{E}_P\left[\tilde{Y}_t\right]$, where $\left(\tilde{Y}_t\right)_{t\geq 0}$ is the counterfactual outcome process of $Y := N^y$ under the intervention that sets treatment to a_0 at all visitation times. Note the different exchangeability condition compared to Ryalen (2024), as Ryalen (2024) expresses exchangeability through the counting process $\mathbb{1}\{\tau^A \leq \cdot\}$; this is actually a weaker condition. Let $\left(T_{(k)}, \Delta_{(k)}, A\left(T_{(k)}\right)\right)$ denote the ordered event times, event types, and treatment decisions. Note that Equation 2 is the same likelihood ratio as in Rytgaard et al. (2022). We also impose the assumption that $N_t := N^y_t + N^{a(\{(0,t] \times \{a_0, a_1\}\})}$ does not explode; we also assume that we work with a version of the compensator such that $\Lambda(\{t\} \times \{y, a\} \times \{a_0, a_1\}) < \infty$ for all $t > 0$. We may generally also work with a compensator Λ that fulfills conditions (10.1.11)-(10.1.13) of Last & Brandt (1995).

NOTE: So the issue is that **Positivity** might not actually hold. If we look at $W(t)$, then it is piecewise constant and only jumps at the treatment times. If it were generally a likelihood ratio, then it would solve Equation 5. For the second term, this implies that $\Lambda^a(\mathrm{d}t) = \sum_k \delta_{T_{(k)}}(\mathrm{d}t)$, so we have placed restrictions on the compensator for N^a. To see this note that,

$$
\begin{aligned}
W(t) &= 1 + \int_0^t W(s-)V(s,m,a)(N(d(s,m,a)) - \Lambda(d(s,m,a)) \\
&= 1 + \int_0^t W(s-)V(s,m,0)(N^{a0}(ds) - \pi_s \Lambda^a(ds)) \\
&\quad\quad + \int_0^t W(s-)V(s,m,1)(N^{a1}(ds) - (1-\pi_s)\Lambda^a(ds)) \\
&= 1 + \int_0^t W(s-)V(s,m,0)(N^{a0}(ds) - \pi_s \Lambda^{a,c}(ds)) \\
&\quad\quad + \int_0^t W(s-)V(s,m,1)(N^{a1}(ds) - (1-\pi_s)\Lambda^{a,c}(ds)) \\
&+ \int_0^t W(s-)V(s,m,0)\big(-\pi_s \Lambda^{a,\text{discrete}}(ds)\big) \\
&+ \int_0^t W(s-)V(s,m,1)\big(-(1-\pi_s)\Lambda^{a,\text{discrete}}(ds)\big).
\end{aligned}
$$

By rearranging, we therefore have that

$$
\begin{aligned}
&W(t) - 1 - \int_0^t W(s-)V(s,m,0)\big(N^{a0}(ds) - \pi_s \Lambda^{a,\text{discrete}}(ds)\big) \\
&- \int_0^t W(s-)V(s,m,1)\big(N^{a1}(ds) - (1-\pi_s)\Lambda^{a,\text{discrete}}(ds)\big) \\
&= - \int_0^t W(s-)(V(s,m,0)\pi_s + V(s,m,1)(1-\pi_s))\lambda^{a,c}(s)\,\mathrm{d}s
\end{aligned}
$$

However, the last integral is only piecewise constant if $\lambda^{a,c} \equiv 0$ because $W(s-)(V(s,m,0)\pi_s + V(s,m,1)(1-\pi_s)) \neq 0$ unless the treatment decision is deterministic. Consequently,

$$
\begin{aligned}
W(t) - 1 - \int_0^t W(s-)V(s,m,0)N^{a0}(ds) \\
- \int_0^t W(s-)V(s,m,1)N^{a1}(ds) \\
= \int_0^t W(s-)((V(s,m,0)\pi_s + V(s,m,1)(1-\pi_s))\Lambda^{a,\text{discrete}}(ds)
\end{aligned}
\tag{1}
$$

In order for Equation 1 to hold true, it must be the case that

$$
\begin{aligned}
&\Delta W(t) - W(t-)V(t,m,0)\Delta N^{a0}(t) - W(t-)V(t,m,1)\Delta N^{a1}(t) \\
&= W(t-)((V(t,m,0)\pi_t + V(t,m,1)(1-\pi_t))\Delta\Lambda^{a,\text{discrete}}(t).
\end{aligned}
$$

However, the left-hand side is only non-zero whenever $\Delta N^a(t) \neq 0$; because, again, $W(t-)((V(t,m,0)\pi_t + V(t,m,1)(1-\pi_t)) \neq 0$, we have must that $\Delta\Lambda^{a,\text{discrete}}(t) = 0$ whenever $\Delta N^a(t) = 0$. Letting $T^a_{(1)}, T^a_{(2)}, \dots$ denote the ordered jump times of N^a, we thus have

$$
\Lambda^a(\mathrm{d}t) = \sum_k A_k(\mathcal{F}_{t-})\delta_{T^a_{(k)}}(\mathrm{d}t), A_k \in (0,1]
$$

Hence

$$N^a(t) - \Lambda^a(t) = \sum_k (1 - A_k) \mathbb{1}\left\{T^a_{(k)} \leq t\right\} \geq 0.$$

Let K be the last number such that $P\left(T^a_{(K)} < \infty\right) > 0$. Note that

$$N^a(t) - \Lambda^a(t) = \sum_{k=1}^{K} (1 - A_k) \mathbb{1}\left\{T^a_{(k)} \leq t\right\}$$

However, the above must also be a zero mean martingale, so that $\mathbb{E}_P\left[(1 - A_k)\mathbb{1}\left\{T^a_{(k)} \leq t\right\}\right] = 0$. Measure theory implies that $(1 - A_k) P\left(T^a_{(k)} \leq t \mid A_k\right) = 0 \quad P$-a.s. for all $t > 0$.

If $P\left(T^a_{(k)} \leq t \mid A_k\right) = 0$ for all $t > 0$ almost surely then $P\left(T^a_{(k)} < \infty \mid A_k\right) = 0$ almost surely and so $T^a_{(k)} = \infty$ almost surely – a contradiction. Therefore $A_k = 1$ for all $k = 1, ..., K$ and N^a is then its own compensator.

Then, the continuous part is zero of the compensator of a. In that case, local independence cannot even motivate this estimand.

- **What about pointwise identification**?
- $\mathbb{E}_P[W(\tau)] = 1$ but not necessarily $\mathbb{E}_P[W(t)] \neq 1$ for all t, so that $W(t)$ is not generally a martingale, then it still be possible to reweight as follows $\mathbb{E}_P\left[\tilde{Y}_\tau\right] = \mathbb{E}_P[Y_\tau W(\tau)]$. Might lose causal interpretation.
- **Might be relevant:** https://pmc.ncbi.nlm.nih.gov/articles/PMC3857358/pdf/nihms529556.pdf

Theorem 0.1: Define

$$\zeta(t, m, a) := \mathbb{1}\{m = y\} + \mathbb{1}\{m = a\} \frac{\mathbb{1}\{a = a_0\}}{\pi_t}$$

If *all* of the following conditions hold:

- **Consistency**: $\tilde{Y}_t \mathbb{1}\left\{\tau^A > \cdot\right\} = Y_t \mathbb{1}\left\{\tau^A > \cdot\right\} \quad P-\text{a.s.}$
- **Exchangeability**: Define $\mathcal{H}_t := \mathcal{F}_t \vee \sigma\left(\tilde{Y}\right)$. The P-$\mathcal{F}_t$ compensator for N^a is also the P-$\mathcal{H}_t$ compensator.
- **Positivity**:

$$W_t := \prod_{j=1}^{N_t} \left(\frac{\mathbb{1}\left\{A\left(T_{(j)}\right) = a_0\right\}}{\pi_{T_{(j)}}\left(\mathcal{F}_{T_{(j-1)}}\right)} \right)^{\mathbb{1}\left\{\Delta_{(j)} = a\right\}} \tag{2}$$

  fulfills that $\int_0^t W(s-) V(s, m, a)(N(d(s, m, a)) - \Lambda(d(s, m, a))$ is a zero mean square-integrable, P-$\mathcal{F}_t$-martingale.

Then,

$$\mathbb{E}_P\left[\tilde{Y}_t\right] = \mathbb{E}_P[Y_t W_t]$$

Proof: We shall use that the likelihood ratio solves a specific stochastic differential equation. To this end, we define the random measure, $\mu(d(t, m, a)) := \zeta(t, m, a)\nu(d(t, m, a))$, where $\nu := \Lambda$. The likelihood ratio process $L(t)$ given in (10.1.14) of Last & Brandt (1995) is defined by

$$
\begin{aligned}
L(t) = \mathbb{1}\{t < T_\infty \wedge T_\infty(\nu)\} L_0 \prod_{n: T_{(n)} \le t} \zeta\big(T_{(n)}, \Delta_{(n)}, A\big(T_{(n)}\big)\big) \\
\prod_{\substack{s \le t \\ N_{s-} = N_s}} \frac{1 - \bar{\nu}\{s\}}{1 - \bar{\mu}\{s\}} \exp\left(\int \mathbb{1}\{s \le t\}(1 - \zeta(s, m, a)) \nu^c(d(s, m, a))\right) \\
+ \mathbb{1}\{t \ge T_\infty \wedge T_\infty(\nu)\} \liminf_{s \to T_\infty \wedge T_\infty(\nu)} L(s).
\end{aligned} \tag{3}
$$

Here $T_\infty := \lim_n T_n$, $T_\infty(\nu) := \inf\{t \ge 0 \mid \nu((0, t] \times \{y, a\} \times \{a_0, a_1\}) = \infty\}$, $\bar{\mu}(\cdot) := \mu(\cdot \times \{y, a\} \times \{a_0, a_1\})$, $\bar{\nu}(\cdot) := \nu(\cdot \times \{y, a\} \times \{a_0, a_1\})$, $\nu^c(\mathrm{d}(s, m, a)) := \mathbb{1}\{\bar{\nu}\{s\} = 0\}\nu(d(s, m, a))$, and $L_0 := W(0) = 1$.

By our assumptions, $T_\infty = \infty \quad P$-a.s. and thus $T_\infty(\nu) = T_\infty = \infty$ in view Theorem 4.1.7 (ii) of Last & Brandt (1995) since $\bar{\nu}\{t\} < \infty$ for all $t > 0$.

Second, note that $\bar{\nu} = \bar{\mu}$. This follows since

$$
\begin{aligned}
\bar{\mu}(A) &= \int_{A \times \{y, a\} \times \{a_0, a_1\}} \zeta(t, m, a) \nu(d(t, m, a)) \\
&= \int_{A \times \{y\} \times \{a_0, a_1\}} \zeta(t, m, a) \nu(d(t, m, a)) + \int_{A \times \{a\} \times \{a_0, a_1\}} \zeta(t, m, a) \nu(d(t, m, a)) \\
&= \int_{A \times \{y\} \times \{a_0, a_1\}} 1 \nu(d(t, m, a)) + \int_{A \times \{a\} \times \{a_0, a_1\}} \left(\frac{\mathbb{1}\{a = a_0\}}{\pi_t}\right) \nu(d(t, m, a)) \\
&= \nu(A \times \{y\} \times \{a_0, a_1\}) + \int_A \Lambda^a(dt) \\
&= \nu(A \times \{y\} \times \{a_0, a_1\}) + \nu(A \times \{a\} \times \{a_0, a_1\}) \\
&= \nu(A \times \{y, a\} \times \{a_0, a_1\}) = \bar{\nu}(A).
\end{aligned}
$$

Thus

$$
\prod_{\substack{s \le t \\ N_{s-} = N_s}} \frac{1 - \bar{\nu}\{s\}}{1 - \bar{\mu}\{s\}} \exp\left(\int \mathbb{1}\{s \le t\}(1 - \zeta(s, m, a)) \nu^c(d(s, m, a))\right) = 1,
$$

and hence

$$
\begin{aligned}
L(t) &= \prod_{n: T_{(n)} \le t} \zeta\big(T_{(n)}, \Delta_{(n)}, A\big(T_{(n)}\big)\big) \\
&\overset{\text{def.}}{=} W(t).
\end{aligned}
$$

Let $V(s, m, a) = \zeta(s, m, a) - 1 + \frac{\bar{\nu}\{s\} - \bar{\mu}\{s\}}{1 - \bar{\mu}\{s\}} = \zeta(s, m, a) - 1$. $L(t)$ will fulfill that

$$
L(t) = L_0 + \int \mathbb{1}\{s \le t\} V(s, m, a) L(s-)[\Phi(d(s, m, a)) - \nu(d(s, m, a))]
$$

if

$$
\begin{aligned}
\mathbb{E}_P[L_0] = 1, \\
\bar{\mu}\{t\} \le 1, \\
\bar{\mu}\{t\} = 1 \text{ if } \bar{\nu}\{t\} = 1, \\
\bar{\mu}[T_\infty \wedge T_\infty(\mu)] = 0 \text{ and } \bar{\nu}[T_\infty \wedge T_\infty(\nu)] = 0.
\end{aligned} \tag{4}
$$

by Theorem 10.2.2 of Last & Brandt (1995). These can be easily verified.

Thus,

$$W(t) = 1 + \int_0^t W(s-)V(s,m,a)(\Phi(d(s,m,a)) - \nu(d(s,m,a)). \tag{5}$$

and it follows that $\int_0^t W(s-)V(s,m,a)(\Phi(d(s,m,a)) - \nu(d(s,m,a))$ is a zero mean$P$-$\mathcal{H}_t$-martingale. From this, we see that $\int_0^t \tilde{Y}_t W_{s-} V(s,m,a)(\Phi(d(s,m,a)) - \nu(d(s,m,a))$ is also a zero mean P-$\mathcal{H}_t$-martingale. This implies that

$$\begin{aligned}
\mathbb{E}_P[Y_t W_t] &\overset{\text{(consistency)}}{=} \mathbb{E}_P\left[\tilde{Y}_t W_t\right] \\
&= \mathbb{E}_P\left[\tilde{Y}_t\right] + \mathbb{E}_P[\int_0^t \tilde{Y}_t W_{s-} V(s,m,a)(\Phi(d(s,m,a)) - \nu(d(s,m,a))] \\
&= \mathbb{E}_P\left[\tilde{Y}_t\right].
\end{aligned}$$

□

Local approach

Theorem 0.2:
- **Consistency**: $\tilde{Y}_\tau \mathbb{1}\{\tau^A > \tau\} = Y_\tau \mathbb{1}\{\tau^A > \tau\} \quad P-\text{a.s.}$
- **Exchangeability**: We have

$$\tilde{Y}_\tau \mathbb{1}\left\{T_{(j)} \leq \tau < T_{(j+1)}\right\} \perp A\left(T_{(k)}\right) \mid \mathcal{F}_{T_{(k-1)}}, T_{(k)}, \Delta_{(k)} = a, \quad \forall j \geq k > 0 \tag{6}$$

- **Positivity**: Let $\pi_{T_{(k)}}\left(\mathcal{F}_{T_{(k-1)}}\right)$ denote the conditional probability of receiving treatment a_0 at time $T_{(k)}$ given $\mathcal{F}_{T_{(k-1)}}$ and $\Delta_{(k)} = a$. We have $\eta > 0$ such that $\pi_{T_{(k)}}\left(\mathcal{F}_{T_{(k-1)}}\right) > \eta \quad P$-a.s. for all $k \geq 1$.

Then the estimand of interest is identifiable, i.e.,

$$\mathbb{E}_P\left[\tilde{Y}_\tau\right] = \mathbb{E}_P[Y_\tau W_\tau]$$

Proof: Write $\tilde{Y}_t = \sum_{k=1}^\infty \mathbb{1}\left\{T_{(k-1)} \leq \tau < T_{(k)}\right\}\tilde{Y}_\tau$. The theorem is shown if we can prove that $\mathbb{E}_P\left[\mathbb{1}\left\{T_{(k-1)} \leq \tau < T_{(k)}\right\}\tilde{Y}_\tau\right] = \mathbb{E}_P\left[\mathbb{1}\left\{T_{(k-1)} \leq \tau < T_{(k)}\right\}Y_\tau W_\tau\right]$ by linearity of expectation. We have that for $k \geq 1$,

$$
\begin{aligned}
&\mathbb{E}_P\left[\mathbb{1}\left\{T_{(k-1)} \leq \tau < T_{(k)}\right\} Y_\tau W(\tau)\right] \\
&= \mathbb{E}_P\left[\mathbb{1}\left\{T_{(k-1)} \leq \tau < T_{(k)}\right\} \mathbb{1}\left\{\tau^A > \tau\right\} Y_\tau W(\tau)\right] \\
&= \mathbb{E}_P\left[\mathbb{1}\left\{T_{(k-1)} \leq \tau < T_{(k)}\right\} \mathbb{1}\left\{\tau^A > \tau\right\} \tilde{Y}_\tau W(\tau)\right] \\
&= \mathbb{E}_P\left[\mathbb{1}\left\{T_{(k-1)} \leq \tau < T_{(k)}\right\} \tilde{Y}_\tau W(\tau)\right] \\
&= \mathbb{E}_P\left[\mathbb{1}\left\{T_{(k-1)} \leq \tau < T_{(k)}\right\} \tilde{Y}_\tau W\left(T_{(k-1)}\right)\right] \\
&= \mathbb{E}_P\left[\mathbb{1}\left\{T_{(k-1)} \leq \tau < T_{(k)}\right\} \tilde{Y}_\tau \left(\frac{\mathbb{1}\left\{A\left(T_{(k-1)}\right) = 1\right\}}{\pi_{T_{(k-1)}}\left(\mathcal{F}_{T_{(k-2)}}\right)}\right)^{\mathbb{1}\left\{\Delta_{(k-1)} = a\right\}} W\left(T_{(k-2)}\right)\right] \\
&= \mathbb{E}_P\Bigg[\mathbb{E}_P\left[\mathbb{1}\left\{T_{(k-1)} \leq \tau < T_{(k)}\right\} \tilde{Y}_\tau \mid \mathcal{F}_{T_{(k-2)}}, \Delta_{(k-1)}, T_{(k-1)}, A\left(T_{(k-1)}\right)\right] \\
&\qquad \times \left(\frac{\mathbb{1}\left\{A\left(T_{(k-1)}\right) = 1\right\}}{\pi_{T_{(k-1)}}\left(\mathcal{F}_{T_{(k-2)}}\right)}\right)^{\mathbb{1}\left\{\Delta_{(k-1)} = a\right\}} W\left(T_{(k-2)}\right)\Bigg] \\
&= \mathbb{E}_P\Bigg[\mathbb{E}_P\left[\mathbb{1}\left\{T_{(k-1)} \leq \tau < T_{(k)}\right\} \tilde{Y}_\tau \mid \mathcal{F}_{T_{(k-2)}}, \Delta_{(k-1)}, T_{(k-1)}\right] \\
&\qquad \times \left(\frac{\mathbb{1}\left\{A\left(T_{(k-1)}\right) = 1\right\}}{\pi_{T_{(k-1)}}\left(\mathcal{F}_{T_{(k-2)}}\right)}\right)^{\mathbb{1}\left\{\Delta_{(k-1)} = a\right\}} W\left(T_{(k-2)}\right)\Bigg] \\
&= \mathbb{E}_P\Bigg[\mathbb{E}_P\left[\mathbb{1}\left\{T_{(k-1)} \leq \tau < T_{(k)}\right\} \tilde{Y}_\tau \mid \mathcal{F}_{T_{(k-2)}}, \Delta_{(k-1)}, T_{(k-1)}\right] \\
&\qquad \times \mathbb{E}_P\left[\left(\frac{\mathbb{1}\left\{A\left(T_{(k-1)}\right) = 1\right\}}{\pi_{T_{(k-1)}}\left(\mathcal{F}_{T_{(k-2)}}\right)}\right)^{\mathbb{1}\left\{\Delta_{(k-1)} = a\right\}} \mid \mathcal{F}_{T_{(k-2)}}, \Delta_{(k-1)}, T_{(k-1)}\right] W\left(T_{(k-2)}\right)\Bigg] \\
&= \mathbb{E}_P\left[\mathbb{E}_P\left[\mathbb{1}\left\{T_{(k-1)} \leq \tau < T_{(k)}\right\} \tilde{Y}_\tau \mid \mathcal{F}_{T_{(k-2)}}, \Delta_{(k-1)}, T_{(k-1)}\right] W\left(T_{(k-2)}\right)\right] \\
&= \mathbb{E}_P\left[\mathbb{E}_P\left[\mathbb{1}\left\{T_{(k-1)} \leq \tau < T_{(k)}\right\} \tilde{Y}_\tau \mid \mathcal{F}_{T_{(k-3)}}, \Delta_{(k-2)}, T_{(k-2)}, A\left(T_{(k-2)}\right)\right] W\left(T_{(k-2)}\right)\right]
\end{aligned}
$$

Iteratively applying the same argument, we get that $\mathbb{E}_P\left[\mathbb{1}\left\{T_{(k-1)} \leq \tau < T_{(k)}\right\} \tilde{Y}_\tau\right] = \mathbb{E}_P\left[\mathbb{1}\left\{T_{(k-1)} \leq \tau < T_{(k)}\right\} Y_\tau W(\tau)\right]$ as needed. □

Bibliography

Last, G., & Brandt, A. (1995). *Marked Point Processes on the Real Line: The Dynamical Approach.* Springer. https://link.springer.com/book/9780387945477

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

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