A note on the potential outcomes framework in continuous time

Johan Sebastian Ohlendorff
johan.ohlendorff@sund.ku.dk
University of Copenhagen

Abstract

In this brief note, we consider the seminal work by Ryalen (2024) and compare it with the approach given in Rytgaard et al. (2022), corresponding to their marked point process settings. We study these works in simple multi-state models.

1 Introduction

We consider a multi-state model with at most one visitation time for the treatment (that is at most one point where treatment may change), no time-varying covariates, and no baseline covariates. In the initial state (0) everyone starts as treated. We consider the setting with no censoring. The multi-state model is shown in Figure 1. We observe the counting processes $N_t = (N_t^{01}, N_t^{02}, N_t^{03}, N_t^{13}, N_t^{23})$ on the canonical filtered probability space $\left(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, P\right)$, where $\mathcal{F}_t = \sigma(N_s \mid s \leq t)$. This means that we can represent the observed data as $O = \left(T_{(1)}, D_{(1)}, A\left(T_{(1)}\right), T_{(2)}\right)$, where $T_{(1)}$ is the first event time, $D_{(1)} \in \{01, 02, 03\}$ is the first event type, $A(T_1) \in \{0, 1\}$ is the treatment at the first event time, and $T_{(2)}$ is the second event time, possibly ∞. We will assume that the distribution of the jump times are continuous and that there are no jumps in common between the counting processes. By a well-known result for marked point processes (Proposition 3.1 of Jacod (1975)), we know there exist functions $h^{ij} : \mathbb{R}_+ \to \mathbb{R}_+$ such that the compensators Λ^{ij} of the counting processes N^{ij} with respect to $P-\mathcal{F}_t$ are given by

$$\Lambda^{0j}(dt) = \mathbb{1}\left\{t \leq T_{(1)}\right\} h^{0j}(t)dt, \quad j = 1, 2, 3$$

$$\Lambda^{i3}(dt) = \mathbb{1}\left\{T_{(1)} < t \leq T_{(2)}\right\} h^{i3}(t)dt, \quad i = 2, 3$$

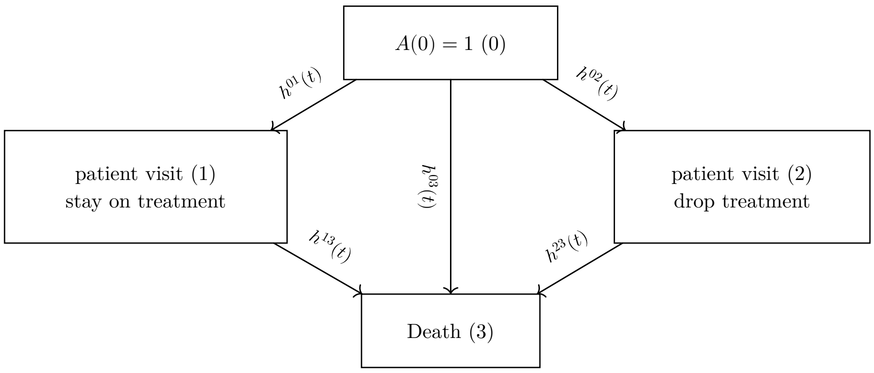

Figure 1: A multi-state model allowing one visitation time for the treatment with the possible treatment values 0/1.

Additionally define

$$\Lambda^a(t) = (h^{01}(t) + h^{02}(t))\mathbb{1}\left\{T_{(1)} \leq t\right\}$$

$$\pi_t(1) = \frac{h^{01}(t)}{h^{01}(t) + h^{02}(t)}\mathbb{1}\left\{T_{(1)} \leq t\right\}$$

Here, we can interpret $\Lambda^a(t)$ as the intensity of the visitation times and $\pi_t(1)$ as the probability of being treated given that you go to the doctor at time t.

2 The potential outcomes framework

To follow along Ryalen (2024), we restrict the observations to the interval $[0, \tau]$ for $\tau > 0$. We first need to define the intervention of interest, defining the counting processes that we would have like to have observed under the intervention. For this define the corresponding "interventional" processes as

$$N_t^{g,0} = 0$$

$$N_t^{g,1} = N_t^{01} + N_t^{02}$$

instead of N_t^{01}, N_t^{02}. This treatment regime defines that the doctor always treats the patient at the visitation time and does not prevent the patient from visiting the doctor if they drop out of the treatment. In contrast, the single "intervention" process

$$N_t^{g^*,0} = 0$$

prevents the patient from visiting the doctor if they drop out of the treatment. The issue in Ryalen (2024) is that we will not be able to differentiate between g and g^* in the likelihood. We let $T^a = \inf_{t>0}\left\{N_t^{g,0} \neq N_t^{01}\right\} \wedge \inf_{t>0}\left\{N_t^{g,1} \neq N_t^{02}\right\} = \inf_{t>0}\left\{N_t^{g,0} \neq 0\right\}$. The outcome of interest is death at time t, i.e.,

$$Y_t = N_t^{13} + N_t^{03} + N_t^{23} = \mathbb{1}\{T_1 \leq t, D_1 = y\} + \mathbb{1}\{T_2 \leq t\}$$

and we want to estimate $\mathbb{E}_P\left[\tilde{Y}_t\right]$ where $\tilde{Y}_t$ denotes the outcome at time t, had the treatment regime (staying on treatment), possibly contrary to fact, been followed.

Theorem 2.1 (Theorem 1 of Ryalen (2024)): We suppose that there exists a potential outcome process $\left(\tilde{Y}_t\right)_{t\in[0,\tau]}$ such that

1. Consistency: $\tilde{Y}_t\mathbb{1}\left\{T^A > t\right\} = Y_t\mathbb{1}\left\{T^A > t\right\}$ for all $t > 0$ P-a.s.
2. Exchangeability: The $P - \mathcal{F}_t$ compensators Λ^{01}, Λ^{02} are also compensators for $\mathcal{G}_t = \mathcal{F}_t \vee \sigma\left(\tilde{Y}_s, \tau \geq s \geq 0\right)$. Here $\tilde{Y}_s$ is added at baseline, so that $\mathcal{G}_0 = \sigma\left(\tilde{Y}_s, \tau \geq s \geq 0\right)$.
3. Positivity: $W_t = \frac{\mathbb{1}\left\{T^A > t\right\}}{\exp\left(-\Lambda_t^{02}\right)} = \frac{1 - \mathbb{1}\left\{T_{(1)} \leq t, D_{(1)} = a, A_{(1)} = 0\right\}}{\exp\left(-\int_0^t \mathbb{1}\left\{s \leq T_{(1)}\right\} h^a(s)\pi_s(0)ds\right)}$ is a uniformly integrable martingale or equivalently that $R^{\text{Pål}}$ given by $dR^{\text{Pål}} = W_\tau dP$ is a probability measure.

Then the estimand of interest is identifiable by

$$\mathbb{E}_P\left[\tilde{Y}_t\right] = \mathbb{E}_P[Y_t W_t] = \mathbb{E}_{R^{\text{Pål}}}[Y_t]$$

We are now in a position, where we can readily compare the approaches in Rytgaard et al. (2022) and Ryalen (2024). Generally speaking the likelihood factorizes as, by the orthogonal martingale assumption,

$$
\begin{aligned}
d\mathrm{P} &= \exp(-\Lambda^{01}(dt) - \Lambda^{02}(dt) - \Lambda^{03}(dt))\left(\Lambda^{01}(dt)\right)^{N^{01}(dt)}\left(\Lambda^{02}(dt)\right)^{N^{02}(dt)}\left(\Lambda^{03}(dt)\right)^{N^{03}(dt)} \\
&\times \exp(-\Lambda^{13}(dt))\left(\Lambda^{13}(dt)\right)^{N^{13}(dt)} \times \exp(-\Lambda^{23}(dt))\left(\Lambda^{23}(dt)\right)^{N^{23}(dt)} \\
&= \exp(-\Lambda^{03}(dt) - \Lambda^{a}(dt))(\pi_t(1)\Lambda^{a}(dt))^{N^{01}(dt)}((1-\pi_t(1))\Lambda^{a}(dt))^{N^{02}(dt)}\left(\Lambda^{03}(dt)\right)^{N^{03}(dt)} \\
&\times \exp(-\Lambda^{13}(dt))\left(\Lambda^{13}(dt)\right)^{N^{13}(dt)} \times \exp(-\Lambda^{23}(dt))\left(\Lambda^{23}(dt)\right)^{N^{23}(dt)} \\
&= \exp(-\Lambda^{03}(dt) - \Lambda^{a}(dt))\left(\Lambda^{03}(dt)\right)^{N^{03}(dt)} \exp(-\Lambda^{13}(dt))\left(\Lambda^{13}(dt)\right)^{N^{13}(dt)} \times \exp(-\Lambda^{23}(dt))\left(\Lambda^{23}(dt)\right)^{N^{23}(dt)} \\
&(\Lambda^{a}(dt))^{N^{01}(dt)}((1-\pi_t(1)))^{N^{02}(dt)} \times (\pi_t(1))^{N^{01}(dt)}((1-\pi_t(1)))^{N^{02}(dt)} \\
&= dQ \times dG
\end{aligned}
$$

where

$$
\begin{aligned}
dQ &= \exp(-\Lambda^{03}(dt))\left(\Lambda^{03}(dt)\right)^{N^{03}(dt)} \exp(-\Lambda^{13}(dt))\left(\Lambda^{13}(dt)\right)^{N^{13}(dt)} \times \exp(-\Lambda^{23}(dt))\left(\Lambda^{23}(dt)\right)^{N^{23}(dt)} \\
&\quad \times \exp(-\Lambda^{a}(dt))(\Lambda^{a}(dt))^{N^{01}(dt)}((1-\pi_t(1)))^{N^{02}(dt)} \\
dG &= (\pi_t(1))^{N^{01}(dt)}((1-\pi_t(1)))^{N^{02}(dt)}
\end{aligned}
$$

Rytgaard et al. (2022) define their target estimand as $\mathbb{E}_{R^{\text{Helene}}}[Y_t]$, where

$$
d\mathrm{R}^{\text{Helene}} = dQ(dt) \times dG^{*}(dt)
$$

where

$$
dG^{*}(dt) = (1)^{N^{01}(dt)}(0)^{N^{02}(dt)}
$$

In contrast, in Ryalen (2024), we have that, by simple multiplication,

$$
\begin{aligned}
d\mathrm{R}^{\text{Pål}} &= W(dt)d\mathrm{P} = \exp(-\Lambda^{03}(dt))\left(\Lambda^{03}(dt)\right)^{N^{03}(dt)} \exp(-\Lambda^{13}(dt))\left(\Lambda^{13}(dt)\right)^{N^{13}(dt)} \times \exp(-\Lambda^{23}(dt))\left(\Lambda^{23}(dt)\right)^{N^{23}(dt)} \\
&\times \exp(-\pi_t(1)\Lambda^{a}(dt))(\pi_t(1)\Lambda^{a}(dt))^{N^{01}(dt)}(0)^{N^{02}(dt)}
\end{aligned}
$$

which does not factorize into Q and G-part of the likelihood. This argument may be made more rigorous by applying Theorem 3 of Ryalen (2024), finding the compensators in the reweighted measure $dR^{\text{Pål}}$.

3 Does the g-formula in Rytgaard et al. (2022) have a causal interpretation?

We now consider the question concerning whether there is a causal interpretation of the g-formula in Rytgaard et al. (2022). A simple result is given in the following theorem. Note that we can also formulated the exchangeability condition for each t separately instead of formulating stochastic process conditions.

Theorem 3.1: We suppose that there exists a potential outcome process $\left(\tilde{Y}_t\right)_{t \in [0,\tau]}$ such that

1. Consistency: $\tilde{Y}_t \mathbb{1}\left\{T^A > t\right\} = Y_t \mathbb{1}\left\{T^A > t\right\}$ for all $t > 0$ P-a.s.
2. Exchangeability: We have
$$\left(\tilde{Y}_t\right)_{t \in [0,\tau]} \perp A\left(T_{(1)}\right) \mid T_{(1)}, D_{(1)}$$
3. Positivity: The measure given by $dR^{\text{Helene}} = W dP$ where $W_t = \left(\frac{\mathbb{1}\left\{A\left(T_{(1)}\right)=1\right\}}{\pi_{T_1}(1)}\right)^{N_t^{01}+N_t^{02}}$ is a probability measure.

Then the estimand of interest is identifiable by
$$\mathbb{E}_P\left[\tilde{Y}_t\right] = \mathbb{E}_P[Y_t W_t] = \mathbb{E}_{R^{\text{Helene}}}[Y_t]$$

Proof: Write $\tilde{Y}_t = \mathbb{1}\left\{t < T_{(1)}\right\}\tilde{Y}_t + \mathbb{1}\left\{T_{(1)} \leq t\right\}\tilde{Y}_t$ Now, we see immediately that
$$\begin{aligned}\mathbb{E}_P\left[\mathbb{1}\left\{t < T_{(1)}\right\}\tilde{Y}_t\right] &= \mathbb{E}_P\left[\mathbb{1}\left\{t < T_{(1)}\right\}\tilde{Y}_t \mathbb{1}\{T^a > t\}\right] \\ &= \mathbb{E}_P\left[\mathbb{1}\left\{t < T_{(1)}\right\}Y_t \mathbb{1}\{T^a > t\}\right] \\ &= \mathbb{E}_P\left[\mathbb{1}\left\{t < T_{(1)}\right\}Y_t\right] \\ &= \mathbb{E}_P\left[\mathbb{1}\left\{t < T_{(1)}\right\}Y_t W_t\right]\end{aligned}$$
since T^a must be $T_{(1)}$ if finite. On the other hand, we have that
$$\begin{aligned}\mathbb{E}_P\left[\mathbb{1}\left\{T_{(1)} \leq t\right\}Y_t W_t\right] &= \mathbb{E}_P\left[\mathbb{1}\left\{T_{(1)} \leq t\right\}\mathbb{1}\{T^a > t\}Y_t W_t\right] \\ &= \mathbb{E}_P\left[\mathbb{1}\left\{T_{(1)} \leq t\right\}\mathbb{1}\{T^a > t\}\tilde{Y}_t W_t\right] \\ &= \mathbb{E}_P\left[\mathbb{1}\left\{T_{(1)} \leq t\right\}\tilde{Y}_t W_t\right] \\ &= \mathbb{E}_P\left[\mathbb{1}\left\{T_{(1)} \leq t\right\}\tilde{Y}_t \left(\frac{\mathbb{1}\left\{A\left(T_{(1)}\right) = 1\right\}}{\pi_{T_{(1)}}(1)}\right)^{\mathbb{1}\left\{D_{(1)}=a\right\}}\right] \\ &= \mathbb{E}_P\left[\mathbb{E}_P\left[\mathbb{1}\left\{T_{(1)} \leq t\right\}\tilde{Y}_t \mid A\left(T_{(1)}\right), D_1, T_1\right]\left(\frac{\mathbb{1}\left\{A\left(T_{(1)}\right) = 1\right\}}{\pi_{T_{(1)}}(1)}\right)^{\mathbb{1}\left\{D_{(1)}=a\right\}}\right] \\ &= \mathbb{E}_P\left[\mathbb{E}_P\left[\mathbb{1}\left\{T_{(1)} \leq t\right\}\tilde{Y}_t \mid D_1, T_1\right]\left(\frac{\mathbb{1}\left\{A\left(T_{(1)}\right) = 1\right\}}{\pi_{T_{(1)}}(1)}\right)^{\mathbb{1}\left\{D_{(1)}=a\right\}}\right] \\ &= \mathbb{E}_P\left[\mathbb{E}_P\left[\mathbb{1}\left\{T_{(1)} \leq t\right\}\tilde{Y}_t \mid D_1, T_1\right]\mathbb{E}_P\left[\left(\frac{\mathbb{1}\left\{A\left(T_{(1)}\right) = 1\right\}}{\pi_{T_{(1)}}(1)}\right)^{\mathbb{1}\left\{D_{(1)}=a\right\}} \mid T_1, D_1\right]\right] \\ &= \mathbb{E}_P\left[\mathbb{E}_P\left[\mathbb{1}\left\{T_{(1)} \leq t\right\}\tilde{Y}_t \mid D_1, T_1\right]\right] \\ &= \mathbb{E}_P\left[\mathbb{1}\left\{T_{(1)} \leq t\right\}\tilde{Y}_t\right]\end{aligned}$$
which suffices to show the claim. □

With more than two events, though, the exchangeability condition becomes more difficult to interpret. In the case with at most three events, for the previous argument to go through, we would need the three exchangeability conditions

$$\left(\tilde{Y}_t \mathbb{1}\left\{T_{(1)} \leq t < T_{(2)}\right\}\right)_{t\in[0,\tau]} \perp A\left(T_{(1)}\right) \mid T_{(1)}, D_{(1)},$$
$$\left(\tilde{Y}_t \mathbb{1}\left\{T_{(2)} \leq t\right\}\right)_{t\in[0,\tau]} \perp A\left(T_{(1)}\right) \mid T_{(1)}, D_{(1)},$$
$$\left(\tilde{Y}_t\right)_{t\in[0,\tau]} \perp A\left(T_{(2)}\right) \mid T_{(2)}, D_{(2)}, A\left(T_{(1)}\right), T_{(1)}, D_{(1)},$$

It would be interesting to see if there are some explicit conditions such that

$$\left(\tilde{Y}_t\right)_{t\in[0,\tau]} \perp A\left(T_{(1)}\right) \mid T_{(1)}, D_{(1)}$$

implies the two first exchangeability conditions. An obvious one is if the event times are independent of the treatment given the history which is unlikely to hold.

The next theorem gives a different causal interpretation of the g-formula in Rytgaard et al. (2022). Unlike the previous theorem, the exchangeability wont have to be specified in terms of $\tilde{Y}_t$ multiplied by a stochastic indicator function if there are more than two events. This issue is however that we are assuming the existence of multiple potential outcome processes and not just one.

Theorem 3.2: We suppose that there exists two potential outcome process $\left(\tilde{Y}_{t,1}\right)_{t\in[0,\tau]}$ and $\left(\tilde{Y}_{t,2}\right)_{t\in[0,\tau]}$ such that these are potential outcomes of $Y_{t,1} = N_t^{03}$ and $Y_{t,2} = N_t^{13} + N_t^{23}$, respectively (the potential outcomes for each possible event where the outcome can occur). Then we obviously define that $\tilde{Y}_t = \tilde{Y}_{t,1} + \tilde{Y}_{t,2}$ and $Y_t = Y_{t,1} + Y_{t,2}$.

1. Consistency: $\tilde{Y}_{t,i} \mathbb{1}\left\{T^A > t\right\} = Y_{t,i} \mathbb{1}\left\{T^A > t\right\}$ for all $t > 0$ P-a.s for $i = 1, 2$.
2. Exchangeability: We have
$$\left(\tilde{Y}_{t,i}\right)_{t\in[0,\tau]} \perp A\left(T_{(1)}\right) \mid T_{(1)}, D_{(1)}$$
for $i = 1, 2$.
3. Positivity: The measure given by $dR^{\text{Helene}} = W dP$ where $W_t = \left(\frac{\mathbb{1}\left\{A\left(T_{(1)}\right)=1\right\}}{\pi_{T_1}(1)}\right)^{N_t^{01}+N_t^{02}}$ is a probability measure.

Then the estimand of interest is identifiable by

$$\mathbb{E}_P\left[\tilde{Y}_t\right] = \mathbb{E}_P[Y_t W_t] = \mathbb{E}_{R^{\text{Helene}}}[Y_t]$$

Proof: Now, we see immediately that

$$\mathbb{E}_P\left[Y_{t,1} W_t\right] = \mathbb{E}_P\left[\tilde{Y}_{t,1}\right]$$

because $\tilde{Y}_{t,1}$ is always $Y_{t,1}$. On the other hand, we have that

$$
\begin{aligned}
\mathbb{E}_P\left[Y_{t,2} W_t\right] &= \mathbb{E}_P\left[Y_{t,2}\left(\frac{\mathbb{1}\left\{A\left(T_{(1)}\right)=1\right\}}{\pi_{T_{(1)}}(1)}\right)^{\mathbb{1}\left\{D_{(1)}=a\right\}}\right] \\
&= \mathbb{E}_P\left[\tilde{Y}_{t,2}\left(\frac{\mathbb{1}\left\{A\left(T_{(1)}\right)=1\right\}}{\pi_{T_{(1)}}(1)}\right)^{\mathbb{1}\left\{D_{(1)}=a\right\}}\right] \\
&= \mathbb{E}_P\left[\mathbb{E}_P\left[\tilde{Y}_{t,2} \mid A\left(T_{(1)}\right), D_1, T_1\right]\left(\frac{\mathbb{1}\left\{A\left(T_{(1)}\right)=1\right\}}{\pi_{T_{(1)}}(1)}\right)^{\mathbb{1}\left\{D_{(1)}=a\right\}}\right] \\
&= \mathbb{E}_P\left[\mathbb{E}_P\left[\tilde{Y}_{t,2} \mid D_1, T_1\right]\left(\frac{\mathbb{1}\left\{A\left(T_{(1)}\right)=1\right\}}{\pi_{T_{(1)}}(1)}\right)^{\mathbb{1}\left\{D_{(1)}=a\right\}}\right] \\
&= \mathbb{E}_P\left[\mathbb{E}_P\left[\tilde{Y}_{t,2} \mid D_1, T_1\right]\mathbb{E}_P\left[\left(\frac{\mathbb{1}\left\{A\left(T_{(1)}\right)=1\right\}}{\pi_{T_{(1)}}(1)}\right)^{\mathbb{1}\left\{D_{(1)}=a\right\}} \mid T_1, D_1\right]\right] \\
&= \mathbb{E}_P\left[\mathbb{E}_P\left[\tilde{Y}_{t,2} \mid D_1, T_1\right]\right] \\
&= \mathbb{E}_P\left[\tilde{Y}_{t,2}\right]
\end{aligned}
$$

which suffices to show the claim. □

Bibliography

Jacod, J. (1975). Multivariate point processes: predictable projection, Radon-Nikodym derivatives, representation of martingales. *Zeitschrift Für Wahrscheinlichkeitstheorie Und Verwandte Gebiete*, *31*(3), 235–253.

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