Supplementary Material for "Causal Mediation Analysis for Longitudinal Mediators and Survival Outcomes"

Proof for Theorem 1

We first provide the required regularity assumptions. (i) Suppose the potential survival time $V_i(z, \mathbf{m})$ as a function of the mediator process $\mathbf{m}$ is Lipschitz continuous on $[0, T]$ with probability one. Namely, there exists a constant $A < \infty$ such that $|V_i(z, \mathbf{m}) - V_i(z, \mathbf{m}')| \le A||m - m'||_2$ for any $z, \mathbf{m}, \mathbf{m}'$ almost surely. (ii) Any path of $\mathbf{m}$ we consider is Lipschitz continuous. There exists a constant B, such that $|m(t_1) - m(t_2)| \le B|t_1 - t_2|$ for any $t_1, t_2 \in [0, T]$.

Fix a time point t and suppose the domain for the covariates is $\mathcal{X}$, with $\mathbf{X}_i^t \in \mathcal{X}$. For any $z, z' \in \{0, 1\}$, we have

$$
\begin{aligned}
&\int_{\mathcal{X}} \int_{R^{[0,t]}} \Pr(V_i > t | Z_i = z, \mathbf{X}_i^t = x_i^t, \mathbf{M}_i^t = \mathbf{m}) dF_{\mathbf{M}_i^t | Z_i = z, \mathbf{X}_i^t = x^t}(\mathbf{m}) dF_{\mathbf{X}_i^t}(x^t) \\
= &\int_{\mathcal{X}} \int_{R^{[0,t]}} \Pr(V_i(z, \mathbf{m}) > t | Z_i = z, \mathbf{X}_i^t = x_i^t, \mathbf{M}_i^t = \mathbf{m}) dF_{\mathbf{M}_i^t | Z_i = z, \mathbf{X}_i^t = x^t}(\mathbf{m}) dF_{\mathbf{X}_i^t}(x^t)
\end{aligned}
$$

For any path $\mathbf{m}$ on the $[0, t]$, we make equal partitions into H pieces at points $\mathcal{M}_H = \{t_0 = 0, t_1 = t/H, t_2 = 2t/H, \cdots, t_H = t\}$ and corresponding values on path $\mathbf{m}$ are $\{m_0, m_1, \cdots, m_H\}$. Then, we consider using a step function from $[0, t] \to \mathcal{R}$ with jumps at points $\mathcal{M}_H$. Denote the step function as $\mathbf{m}_H$, which is:

$$
\mathbf{m}_H(x) = \begin{cases} \mathbf{m}(0) = m_0 & 0 \le x < t/H, \\ \mathbf{m}(t/H) = m_1 & t/H \le x < 2t/H, \\ \cdots & \\ \mathbf{m}((H-1)t/H) = m_H & (H-1)t/H \le x \le t. \end{cases}
$$

We employ this step function $\mathbf{m}_H(x)$ to approximate function $\mathbf{m}$. First, given $\mathbf{m}$ is Lipschitz continuous, there exists $B > 0$ such that $|m(x_1) - m(x_2)| \le B|x_1 - x_2|$. Therefore, the step function $\mathbf{m}_H$ can approximate the original function $\mathbf{m}$ well as H goes up,

$$
||\mathbf{m}_H - \mathbf{m}||_2 \le \sum_{i=1}^{H} \frac{t}{H} B^2 \frac{t^2}{H^2} \asymp O(H^{-2}).
$$

Therefore, we can approximate the survival probability given a continuous mediator process with the mediator values on the jumps, $(m_0, m_1, \cdots, m_H)$. That is,

$$\int_{R^{[0,t]}} \Pr(V_i(z,\mathbf{m}) > t | Z_i = z, \mathbf{X}_i^t = \mathbf{x}^t, \mathbf{M}_i^t = \mathbf{m}) \times \mathrm{d}\{\mathrm{F}_{\mathbf{M}_i^t|Z_i=z',\mathbf{X}_i^t=\mathbf{x}^t}(\mathbf{m})\}$$
$$\asymp \int_{R^{[0,t]}} \Pr(V_i(z,\mathbf{m}_H) > t | Z_i = z, \mathbf{X}_i^t = \mathbf{x}^t, \mathbf{M}_i^t = \mathbf{m}_H) \times \mathrm{d}\{\mathrm{F}_{\mathbf{M}_i^t|Z_i=z',\mathbf{X}_i^t=\mathbf{x}^t}(\mathbf{m}_H)\} + O(H^{-2}).$$

This step applies the regularity condition that the potential survival time $V_i^t(z,\mathbf{m})$ as a function of $\mathbf{m}$ is continuous with the L_2 metrics of $\mathbf{m}$. As the values of steps function $\mathbf{m}_H$ are completely determined by the values on finite jumps, we can further reduce the conditional survival probability to,

$$\asymp \int_{R^H} E(Y_i^t(z,\mathbf{m}_H) | Z_i = z, \mathbf{X}_i^t = \mathbf{x}^t, m_0, m_1, m_2, \cdots m_H)$$
$$\times \mathrm{d}\{\mathrm{F}_{m_0,m_1,\cdots,m_H|Z_i=z',\mathbf{X}_i^t=\mathbf{x}^t}(m_0, m_1, m_2, \cdots m_H)\} + O(H^{-2}).$$

Under assumption 1, we have,

$$\begin{aligned}
&\mathrm{d}\{\mathrm{F}_{m_0,m_1,\cdots,m_H|Z_i=z',\mathbf{X}_i^t=\mathbf{x}^t}(m_0, m_1, m_2, \cdots m_H)\} \\
&= \mathrm{d}\{\mathrm{F}_{m_0(z'),m_1(z'),\cdots,m_H(z')|\mathbf{X}_i^t=\mathbf{x}^t}(m_0, m_1, m_2, \cdots m_H)\}, \\
&= \mathrm{d}\{\mathrm{F}_{\mathbf{m}_H(z')|\mathbf{X}_i^t=\mathbf{x}^t}(\mathbf{m}_H)\}.
\end{aligned}$$

With a slightly abuse of notations, let $\mathbf{m}_H(z)$ denote the potential step functions induced by the original potential process $\mathbf{M}_i^t(z)$ and $m_i(z)$ to denote potential values of $\mathbf{M}_i^t(z)$ evaluated at point $t_i = it/H$. Under Assumption 2, we can choose a large H such that $t/H \le \varepsilon$. Then we have the following conditional independence conditions,

$$\begin{aligned}
&V_i(z,\mathbf{m}_H) \perp\!\!\!\perp m_0 | Z_i, \mathbf{X}_i^t, \\
&V_i(z,\mathbf{m}_H) \perp\!\!\!\perp (m_1 - m_0) | Z_i, \mathbf{X}_i^t, \mathbf{m}_H^0, \\
&V_i(z,\mathbf{m}_H) \perp\!\!\!\perp (m_2 - m_1) | Z_i, \mathbf{X}_i^t, \mathbf{m}_H^{t/H}, \\
&\qquad\qquad \cdots \\
&V_i(z,\mathbf{m}_H) \perp\!\!\!\perp (m_H - m_{H-1}) | Z_i, \mathbf{X}_i^t, \mathbf{m}_H^{t(H-1)/H},
\end{aligned}$$

where are equivalent to,

$$\begin{aligned}
&V_i(z,\mathbf{m}_H) \perp\!\!\!\perp m_0 | Z_i, \mathbf{X}_i^t, \\
&V_i(z,\mathbf{m}_H) \perp\!\!\!\perp (m_1 - m_0) | Z_i, \mathbf{X}_i^t, m_0, \\
&V_i(z,\mathbf{m}_H) \perp\!\!\!\perp (m_2 - m_1) | Z_i, \mathbf{X}_i^t, m_0, m_1, \\
&\qquad\qquad \cdots \\
&V_i(z,\mathbf{m}_H) \perp\!\!\!\perp (m_H - m_{H-1}) | Z_i, \mathbf{X}_i^t, m_0, m_1 \cdots, m_{H-1},
\end{aligned}$$

as the step function $m_H^{it/H}, i \leq H$ is completely determined by values at jumps $\{m_0, \cdots, m_i\}$. With the established conditional independence, we have,

$$\Pr(V_i(z, \mathbf{m}_H) > t | Z_i = z, \mathbf{X}_i^t = \mathbf{x}^t, m_0, m_1, m_2, \cdots m_H) = \Pr(V_i(z, \mathbf{m}_H) > t | Z_i = z, \mathbf{X}_i^t = \mathbf{x}^t).$$

With similar arguments, we can show that,

$$\begin{aligned}
&\Pr(V_i(z, \mathbf{m}_H) > t | Z_i = z, \mathbf{X}_i^t = \mathbf{x}^t) = \Pr(V_i(z, \mathbf{m}_H) > t | Z_i = z', \mathbf{X}_i^t = \mathbf{x}^t), \\
&= \Pr(V_i(z, \mathbf{m}_H) > t | Z_i = z, \mathbf{X}_i^t = \mathbf{x}^t, m_0 = m_0(z'), \cdots m_H = m_H(z')), \\
&= \Pr(V_i(z, \mathbf{m}_H) > t | Z_i = z, \mathbf{X}_i^t = \mathbf{x}^t, \mathbf{m}_H(z') = \mathbf{m}_H), \\
&= \Pr(V_i(z, \mathbf{m}_H) > t | \mathbf{X}_i^t = \mathbf{x}^t, \mathbf{m}_H(z') = \mathbf{m}_H).
\end{aligned}$$

As a conclusion, we have shown that,

$$\begin{aligned}
&\int_{\mathcal{X}} \int_{R^{[0,t]}} \Pr(V_i(z, \mathbf{m}) > t | Z_i = z, \mathbf{X}_i^t = \mathbf{x}^t, \mathbf{M}_i^t = \mathbf{m}) dF_{\mathbf{X}_i^t}(x^t) \times \mathrm{d}\{\mathrm{F}_{\mathbf{M}_i^t | Z_i = z', \mathbf{X}_i^t = \mathbf{x}^t}(\mathbf{m})\}, \\
&\asymp \int_{\mathcal{X}} \int_{R^{[0,t]}} \Pr(V_i(z, \mathbf{m}_H) > t | \mathbf{X}_i^t = \mathbf{x}^t, \mathbf{m}_H(z') = \mathbf{m}_H) \times \mathrm{d}\{\mathrm{F}_{\mathbf{m}_H(z') | \mathbf{X}_i^t = \mathbf{x}^t}(\mathbf{m}_H)\} dF_{\mathbf{X}_i^t}(x^t) + O(H^{-2}), \\
&\asymp \int_{\mathcal{X}} \Pr(V_i(z, \mathbf{m}_H(z')) > t | \mathbf{X}_i^t = \mathbf{x}^t) + O(H^{-2}) \asymp \int_{\mathcal{X}} \Pr(V_i(z, \mathbf{m}(z')) > t | \mathbf{X}_i^t = \mathbf{x}^t) + O(H^{-2}).
\end{aligned}$$

The last equivalence follows from the regularity condition of $V_i(z, \mathbf{m}(z'))$ as a function of $\mathbf{m}(z')$. Let H goes to infinity, we have,

$$\begin{aligned}
&\int_{\mathcal{X}} \int_{R^{[0,t]}} \Pr(V_i > t | Z_i = z, \mathbf{X}_i^t = \mathbf{x}^t, \mathbf{M}_i^t = \mathbf{m}) \mathrm{dF}_{\mathbf{X}_i^t}(\mathbf{x}^t) \times \mathrm{d}\{\mathrm{F}_{\mathbf{M}_i^t | Z_i = z', \mathbf{X}_i^t = \mathbf{x}^t}(\mathbf{m})\} \\
&= \int_{\mathcal{X}} \Pr(V_i(z, \mathbf{m}(z')) > t | \mathbf{X}_i^t = \mathbf{x}^t) \mathrm{dF}_{\mathbf{X}_i^t}(\mathbf{x}^t) = \Pr(V_i(z, \mathbf{m}(z')) > t) = S_{z,z'}(t)
\end{aligned}$$

Under Assumption 3, the conditional survival function can be estimated with a non-parametric Kaplan Meier estimator,

$$\begin{aligned}
\Pr(V_i > t | Z_i = z, \mathbf{X}_i^t = \mathbf{x}^t, \mathbf{M}_i^t = \mathbf{m}) &= \prod_{k=1}^{K} \Pr(\tilde{V}_i > t_k | \tilde{V}_i > t_{k-1}, \mathbf{M}_i^t = m.\mathbf{X}_i^t, Z_i = z) \\
&= \prod_{k=1}^{K} \Pr(\tilde{V}_i > t_k | \tilde{V}_i > t_{k-1}, \mathbf{M}_i = m^{t_k}.\mathbf{X}_i^{t_k}, Z_i = z),
\end{aligned}$$

where $t_1 < t_2 <, \cdots, < t_k <, \cdots, < t_K < \cdots$ is the time grid where we observe failure event ($\delta_i = 1$) and the selected fixed time point t lies between t_K and t_{K+1}. Hence, we complete the proof for Theorem 1.

Details of Gibbs Sampler

In this section, we provide detailed descriptions on the Gibbs sampler for the model in Section 4. The sampling of mediator process is similar to the one in Kowal and Bourgeois (2020) and Zeng et al. (2021a). Therefore, we omit the details for simplicity and refer the reader to Zeng et al. (2021b).

Next, we describe the sampling for the survival model. As we have the model,

$$\lambda(t) = \lambda_0(t) \exp(Z_i\alpha + X_{ij}'\beta_S + f\{\mathbf{M}_i^t, \gamma\}).$$

The survival function for a specific subject becomes,

$$S_i(t) = \Pr(V_i > t) = \exp(-H_i(t)) = \exp(-\int_0^t h_i(s)ds)$$

where $H_i(t)$ is the cumulative hazard function, which is a right-continuous increasing function with $H_i(0) = 0$. For a given observation $(\tilde{V}_i, \delta_i, \mathbf{X}_{ij}, M_{ij}, Z_i)$, the likelihood of this observation is,

$$L_{ij} = (1 - \delta_{ij}) \Pr(V_i > t_{ij}) + \delta_{ij} \Pr(V_i = t_{ij}).$$

where δ_{ij} is the indicator for whether the subject is still alive at time point t_{ij}. To derive an explicit formula for the likelihood, we let $\lambda_1, \lambda_2, \cdots, \lambda_K$ be the baseline hazard for a specified time grids $t_1 < t_2, \cdots, < t_K$ that at least one failure happens in each bin $(t_{k-1}, t_k]$. Then the cumulative hazard function becomes,

$$H_i(t) = \sum_{k=1}^{K} \lambda_k U_{i,k}(t, \alpha, \beta_S, \gamma),$$

where

$$\begin{aligned} U_{i,k}(t, \alpha, \beta_S, \gamma) &= (t_k - t_{k-1}) \exp(Z_i\alpha + X_i^{t_k}\beta_S + f\{\mathbf{M}_i^{t_k}, \gamma\}) \ \text{ if } t > t_k \\ U_{i,k}(t, \alpha, \beta_S, \gamma) &= (t - t_{k-1}) \exp(Z_i\alpha + X_i^{t_{k-1}}\beta_S + f\{\mathbf{M}_i^{t}, \gamma\}) \ \text{ if } t < t_k. \end{aligned}$$

As such, we can express the likelihood of the data as a function of parameter $(\{\lambda_k\}_{k=1}^K, \alpha, \beta_S, \gamma)$. First, we describe the prior of baseline hazard rate $\{\lambda_k\}_{k=1}^K$. We specify a Gamma Process prior on λ_k, that is the increments are independent across each other and follow a Gamma distribution, $\lambda_k \sim Gamma(\alpha_k, \beta_k)$. We specify α_k, β_k in the following way that, we let $\alpha_k = A\alpha(t_k)$ and $\beta_k = B$, where $\alpha(t)$ is strictly increasing function that captures the mean of the hazard rate. For example, when $\alpha(t) = t$, then $E\{\lambda_k\} = t_k A/B$. For the hyperparameter A, B, we specify a Gamma prior such that $A, B \sim Gamma(\varepsilon, \varepsilon)$ with $\varepsilon = 0.001$.

Then the conditional posterior distributions of the parameters in the sample are:

- Baseline hazard rate $\lambda_k|-$:

$$p(\lambda_k|-) = Ga(\lambda_k|\alpha_k + n_k, \beta_k + m_k(\alpha, \beta_S, \gamma)),$$

where n_k is the number of failure in $(t_{k-1}, t_k]$ and $m_k(\alpha, \beta_S, \gamma) = \sum_{i=1}^{N} \sum_{j=1}^{n_i} U_{i,k}(t_{ij}, \alpha, \beta_S, \gamma)$.

- The hyperparameter for the Gamma Process A, B—-,

$$p(A|-) \propto A^{\varepsilon-1} \exp(-\varepsilon A) B^{\varepsilon\alpha(t_K)} \prod_{k=1}^{K} \frac{\lambda_k^{A(\alpha(t_k)-\alpha(t_{k-1}))}}{\Gamma(A(\alpha(t_k) - \alpha(t_{k-1}))},$$
$$p(B|-) \propto B^{A\alpha(t)+\varepsilon-1} \exp(-B(\varepsilon + \sum_{k=1}^{K} \lambda_k)$$

This step can be updated using a one step of Metropolis random walk.

- The coefficient for treatment, covariates and mediator process: $\alpha, \beta_S, \gamma|-$

$$p(\alpha, \beta_S, \gamma|-) \propto p(\alpha, \beta_S, \gamma) \exp(\sum_{i=1}^{N} \delta_i(Z_i\alpha + X_{in_i}\beta_S + f\{\mathbf{M}_i^{t_{in_i}}, \gamma) - \sum_{k=1}^{K} \lambda_k m_k(\alpha, \beta_S, \gamma)),$$

This step can be updated efficiently using the adaptive rejection methods in Gilks and Wild (1992) as the density is log-concave in $(\alpha, \beta_S, \gamma)$. The parameterization of the cumulative model $f\{\mathbf{M}_i^t, \gamma)$ is similar in the construction of spline basis in mediator process. We refer the readers to Kowal and Bourgeois (2020) and Zeng et al. (2021a) for details.

References

Gilks, W. R. and Wild, P. (1992), "Adaptive rejection sampling for Gibbs sampling," *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 41, 337–348.

Kowal, D. R. and Bourgeois, D. C. (2020), "Bayesian Function-on-Scalars Regression for High-Dimensional Data," *Journal of Computational and Graphical Statistics*, 29, 629–638.

Zeng, S., Rosenbaum, S., Archie, E., Alberts, S., and Li, F. (2021a), "Causal mediation analysis for sparse and irregular longitudinal data," *Annals of Applied Statistics*, 15, 747–767.

Zeng, S., Stacy, R., Susan, A., Elizabeth, A., and Li, F. (2021b), "Supplement to "Causal mediation analysis for sparse and irregular longitudinal data."," .