Stabilized Neural Prediction of Potential Outcomes in Continuous Time

Konstantin Hess[1,2], Stefan Feuerriegel[1,2]

1) Institute of AI in Management, LMU Munich

2) Munich Center for Machine Learning (MCML)

Motivation

In medicine, there is a growing interest in estimating treatment effects / potential outcomes from patient trajectories using observational data.

Methods should fulfil two requirements:

Continuous time modeling
- Patient data may be recorded irregularly.
- Patients may need immediate treatments.

➔ Method need to model patient trajectories not in discrete time (e.g., fixed daily / hourly timestamps) but in **continuous time** (i.e., actual timestamps).

Adjustments for time-varying confounding
- Future patient covariates may affect treatment decisions in observational data.
- These time-varying confounders may lead to bias as they are unobserved during inference time.

➔ Methods need to **adjust for time-varying confounding**, as they would otherwise target an incorrect estimand.

Related work

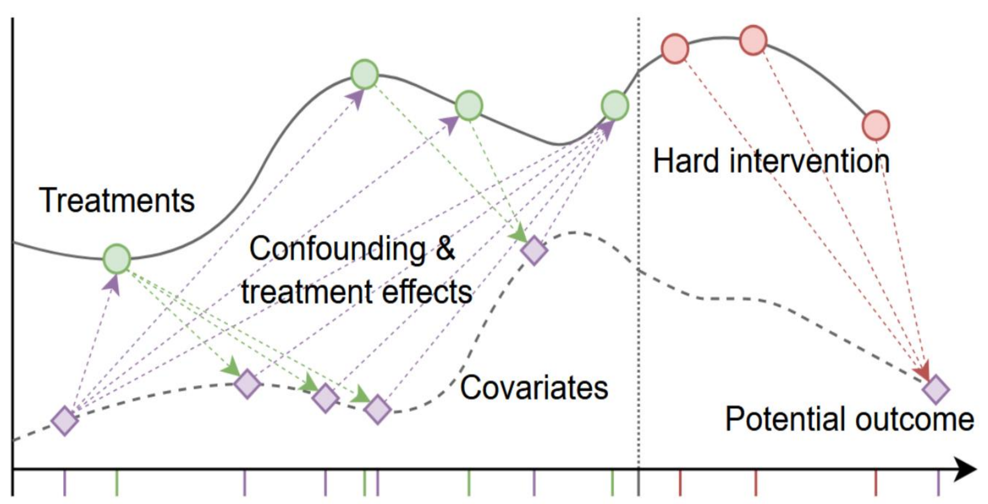
Neural methods for potential outcome estimation over time

	Correct timestamps?	Adjustment for time-varying confounding?	Existing works
(1) Neural methods in discrete time	✗ ✗	✗ ✓	**CRN** (Bica et al., 2020), **CT** (Melnychuk et al., 2022) **RMSNs** (Lim et al., 2018), **G-Net** (Li et al., 2021)
(2) Neural methods in continuous time	✓	✗	**TE-CDE** (Seedat et al., 2022)
SCIP-Net (ours)	✓	✓	—

- Existing neural methods primarily focus on **discrete time**.
- No method properly adjusts for time-varying confounding in continuous time.
- Our **SCIP-Net** (*stabilized continuous-time inverse propensity network*) is the only method that adjust for time-varying confounding in continuous time.

Our approach

- We derive **tractable** inverse-propensity weighting in continuous time via geometric product-integration.
- We **reduce estimation variance** via continuous-time scaling weights.



- We develop a **tailored neural method**, the SCIP-Net.

Problem Formulation

Observational likelihood:

$$\mathrm{d}\mathbb{P}_0(\bar{H}_\tau) = \mu_{0,0}(X_0) \prod_{s\in(0,\tau]} \Big(\big(\mathrm{d}\Lambda_0^x(s \mid \bar{H}_{s-})\mu_{0,s}(X_s \mid \bar{H}_{s-})\big)^{N^x(\mathrm{d}s)} \big(1-\mathrm{d}\Lambda_0^x(s \mid \bar{H}_{s-})\big)^{1-N^x(\mathrm{d}s)} \times \big(\mathrm{d}\Lambda_0^a(s \mid \bar{H}_{s-})\pi_{0,s}(A_s \mid \bar{H}_{s-})\big)^{N^a(\mathrm{d}s)} \big(1-\mathrm{d}\Lambda_0^a(s \mid \bar{H}_{s-})\big)^{1-N^a(\mathrm{d}s)} \Big)$$

Estimation task: $\mathbb{E}\Big[Y_\tau[\underline{a}_{*,t}, \underline{n}^a_{*,t}] \mid \bar{H}_{t-} = \bar{h}_{t-}\Big]$

Identifiability: (1) Consistency, (2) overlap, (3) sequential ignorability

Interventional distribution:

$$\mathrm{d}\mathbb{P}_*(\bar{H}_\tau) = \mathrm{d}\mathbb{P}_{Q_0,G_*}(\bar{H}_\tau) = \mu_{0,0}(X_0) \prod_{s\in(0,\tau]} \mathrm{d}G_{*,s}(\bar{H}_s)\, \mathrm{d}Q_{0,s}(\bar{H}_s)$$

where $\mathrm{d}G_{*,s}(\bar{H}_s) = \big(\mathrm{d}\Lambda_*^a(s \mid \bar{H}_{s-})\, \pi_{*,s}(A_s \mid \bar{H}_{s-})\big)^{N^a(\mathrm{d}s)} \big(1-\mathrm{d}\Lambda_*^a(s \mid \bar{H}_{s-})\big)^{1-N^a(\mathrm{d}s)}$

is the interventional part (i.e., treatments and treatment times) and

$$\mathrm{d}Q_{0,s}(\bar{H}_s) = \big(\mathrm{d}\Lambda_0^x(s \mid \bar{H}_{s-})\mu_{0,s}(X_s \mid \bar{H}_{s-})\big)^{N^x(\mathrm{d}s)} \big(1-\mathrm{d}\Lambda_0^x(s \mid \bar{H}_{s-})\big)^{1-N^x(\mathrm{d}s)}$$

is the observational part (i.e., patient covariates).

Proposition 1 (continuous time IPW): Under assumptions (1)-(3), we can estimate the potential outcome via

$$\mathbb{E}\Big[Y_\tau[\underline{a}_{*,t}, \underline{n}^a_{*,t}] \,\Big|\, \bar{H}_{t-} = \bar{h}_{t-}\Big] = \mathbb{E}\Big[Y_\tau \prod_{s\geq t} W_s \,\Big|\, (\underline{A}_t, \underline{N}^a_t) = (\underline{a}_{*,t}, \underline{n}^a_{*,}), \bar{H}_{t-} = \bar{h}_{t-}\Big]$$

where the inverse propensity weights are defined as $W_s \equiv w_s(\bar{H}_s) = \dfrac{\mathrm{d}G_{*,s}(\bar{H}_s)}{\mathrm{d}G_{0,s}(\bar{H}_s)}$.

SCIP-Net

Objective: Find the optimal parameters $\hat{\phi}$ of a neural network m_ϕ via

$$\hat{\phi} = \arg\min_\phi \mathbb{E}_{\mathbb{P}_*}\Big[\Big(Y_\tau[\underline{a}_{*,t}, \underline{n}^a_{*,t}] - m_\phi(\underline{A}_t, \underline{N}^a_t, \bar{H}_{t-})\Big)^2 \,\Big|\, \bar{H}_{t-} = \bar{h}_{t-}\Big]$$
$$= \arg\min_\phi \mathbb{E}_{\mathbb{P}_0}\Big[\Big(Y_\tau - m_\phi(\underline{A}_t, \underline{N}^a_t, \bar{H}_{t-})\Big)^2 \prod_{s\geq t} W_s \,\Big|\, (\underline{A}_t, \underline{N}^a_t) = (\underline{a}_{*,t}, \underline{n}^a_{*,t}), \bar{H}_{t-} = \bar{h}_{t-}\Big].$$

Proposition 2 (tractable unstaibilized weights): The unstabilized inverse propensity weights in Proposition 1 satisfy

$$\prod_{s\geq t} W_s \,\Big|\, \Big((\underline{A}_t, \underline{N}^a_t) = (\underline{a}_{*,t}, \underline{n}^a_{*,t}), \bar{H}_{t-} = \bar{h}_{t-}\Big) = \prod_{j=1}^{J} W_{t^a_{*,j}} \,\Big|\, \Big((\underline{A}_t, \underline{N}^a_t) = (\underline{a}_{*,t}, \underline{n}^a_{*,t}), \bar{H}_{t-} = \bar{h}_{t-}\Big)$$

where $W_{t^a_{*,j}} = \dfrac{\exp \int_{s\in[t^a_{*,j-1}, t^a_{*,j})} \lambda_0^a(s \mid \bar{H}_{s-})\, \mathrm{d}s}{\lambda_0^a(t^a_{*,j} \mid \bar{H}_{t^a_{*,j}-})\, \pi_{0,t^a_{*,j}}(a_{*,t^a_{*,j}} \mid \bar{H}_{t^a_{*,j}-})}$.

Definition 1 (stabilized weights): Let the scaling factor Ξ_s be given by

$$\Xi_s \equiv \xi_s(\bar{A}_s, \bar{N}^a_s) = \frac{\mathrm{d}G_{0,s}(\bar{A}_s, \bar{N}^a_s)}{\mathrm{d}G_{*,s}(\bar{A}_s, \bar{N}^a_s)}.$$

Define the **stabilized weights** $\widetilde{W}_s$ as $\widetilde{W}_s = \Xi_s W_s$.

Proposition 3 (optimal parameters): The optimal parameters are equivalently given by

$$\hat{\phi} = \arg\min_\phi \mathbb{E}_{\mathbb{P}_0}\Big[\Big(Y_\tau - m_\phi(\underline{A}_t, \underline{N}^a_t, \bar{H}_{t-})\Big)^2 \prod_{s>t} \widetilde{W}_s \,\Big|\, (\underline{A}_t, \underline{N}^a_t) = (\underline{a}_{*,t}, \underline{n}^a_{*,t}), \bar{H}_{t-} = \bar{h}_{t-}\Big].$$

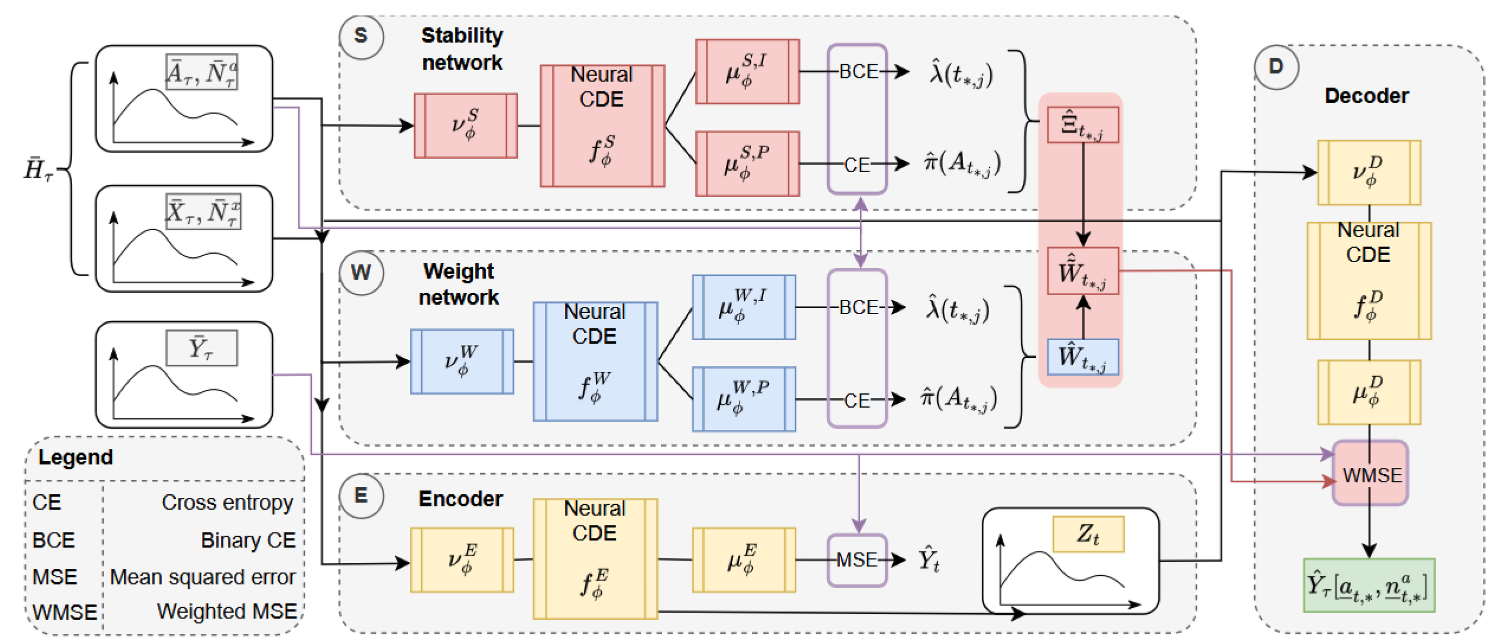
Proposition 4 (tractable staibilized weights): The stabilized scaling factor Ξ_s from Definition 1 and Proposition satisfies

$$\prod_{s\geq t} \Xi_s \,\Big|\, \Big((\underline{A}_t, \underline{N}^a_t) = (\underline{a}_{*,t}, \underline{n}^a_{*,t}), (\bar{A}_{t-}, \bar{N}^a_{t-}) = (\bar{a}_{t-}, \bar{n}^a_{t-})\Big)$$
$$= \prod_{j=1}^{J} \Xi_{t^a_{*,j}} \,\Big|\, \Big((\underline{A}_t, \underline{N}^a_t) = (\underline{a}_{*,t}, \underline{n}^a_{*,t}), (\bar{A}_{t-}, \bar{N}^a_{t-}) = (\bar{a}_{t-}, \bar{n}^a_{t-})\Big)$$

where $\Xi_{t^a_{*,j}} = \dfrac{\lambda_0^a(t^a_{*,j} \mid \bar{A}_{t^a_{*,j}-}, \bar{N}^a_{t^a_{*,j}-})\, \pi_{0,t^a_{*,j}}(a_{t^a_{*,j}} \mid \bar{A}_{t^a_{*,j}-}, \bar{N}^a_{t^a_{*,j}-})}{\exp \int_{s\in[t^a_{*,j-1}, t^a_{*,j})} \lambda_0^a(s \mid \bar{A}_{s-}, \bar{N}^a_{s-})\, \mathrm{d}s}$.

➔ Together, Propositions 2, 3 and 4 yield a **(i) tractable objective function** that relies on **(ii) stabilized inverse propensity weights**.

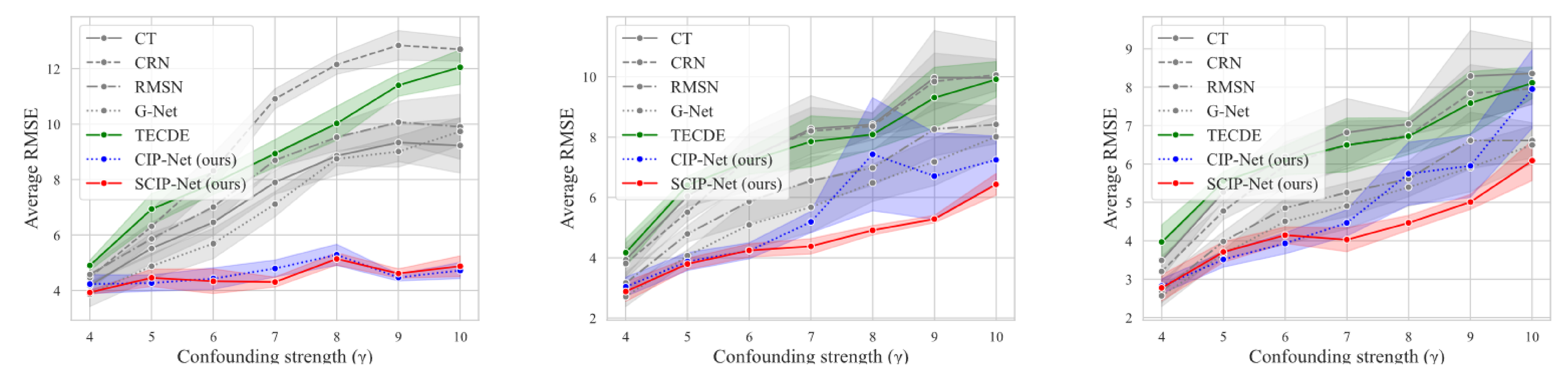
Neural architecture of SCIP-Net



- First, an **encoder** network encodes the patient history.
- **Sub-networks** estimate **scaling factors** and **unstabilized weights**.
- The encoded history is passed to the decoder.
- The **decoder** is trained on the interventional treatment sequence via stabilized inverse propensity weights.

Results

Tumor growth data:



(a) One day ahead prediction (b) Two days ahead prediction (c) Three days ahead prediction

MIMIC-III data:

Prediction window	CT	CRN	RMSNs	G-Net	TE-CDE	**CIP-Net** (ours)	**SCIP-Net** (ours)	**Rel. improvement**
$(\tau - t) = 1$ hours	1.052 ± 0.069	1.049 ± 0.065	1.075 ± 0.074	1.021 ± 0.069	0.915 ± 0.025	$\mathbf{0.876 \pm 0.041}$	0.877 ± 0.044	+4.1%
$(\tau - t) = 2$ hours	1.196 ± 0.272	1.088 ± 0.374	1.130 ± 0.274	1.095 ± 0.335	0.784 ± 0.145	0.785 ± 0.117	$\mathbf{0.634 \pm 0.148}$	+19.1%
$(\tau - t) = 3$ hours	1.444 ± 0.232	1.262 ± 0.355	1.300 ± 0.304	1.330 ± 0.198	1.240 ± 0.242	1.291 ± 0.400	$\mathbf{1.089 \pm 0.322}$	+12.2%