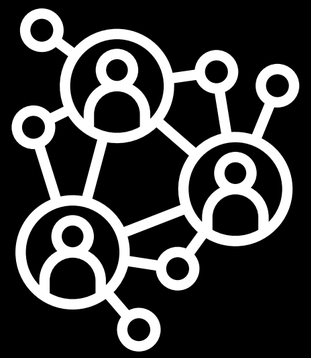
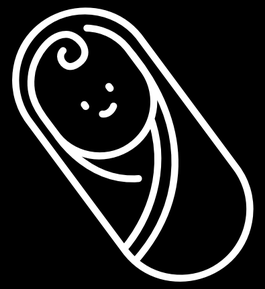
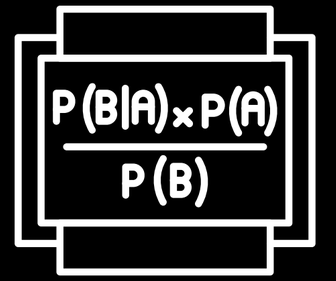




Learning DBNs & Non-Homogeneous DBNs from Incomplete data in Intensive Care

Federico Pirola, Fabio Stella, Marco Grzegorczyk, Anna Lavizzari



UNIVERSITÀ DEGLI STUDI DI MILANO BICOCCA
MAD Lab
DISCO DIPARTIMENTO DI INFORMATICA SISTEMISTICA E COMUNICAZIONE
university of groningen
Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico
Anthem

Sections

PRELIMINARIES

- Dynamic Bayesian Networks: definition and application domains
- Bayesian Statistics and Montecarlo methods

LEARNING DBNs FROM INCOMPLETE DATA

- Issues in Observational Settings
- LUME-DBN: A full Bayesian method for learning DBNs from Incomplete Data
- Applications in Intensive Care
  - Physionet Case Study
  - Bronchopulmonary Dysplasia Case Study

NON STATIONARY TEMPORAL MODELS

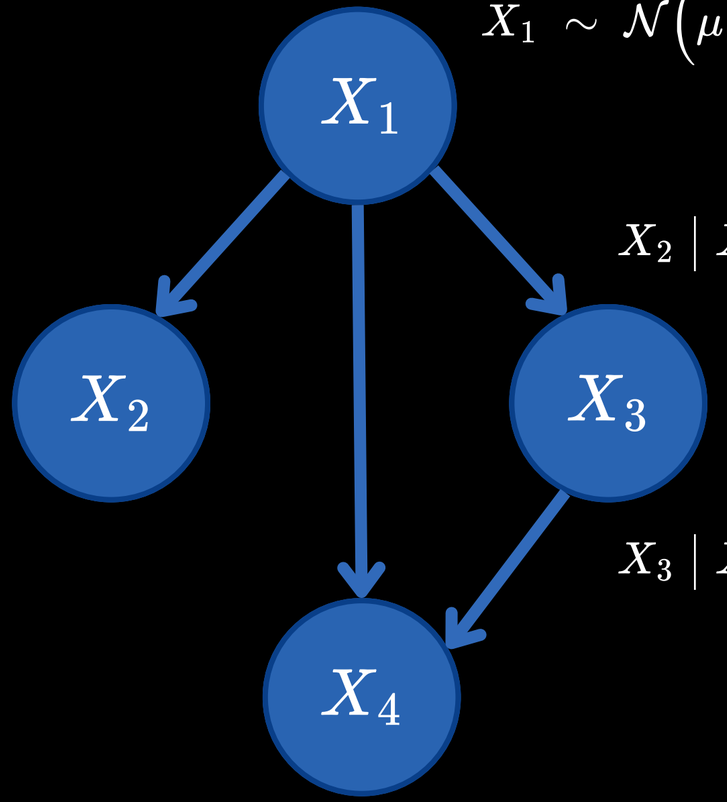
- Non-stationary and Non-Homogeneous DBNs
- LUME-NH: A full Bayesian method for learning NH-DBNs from Incomplete Data

Dynamic Bayesian Networks

Probabilistic Graphical Models

Highly interpretable Machine Learning models widely employed in biomedical domains

- **MODELS:** modeling the dependencies between multiple random variables
- **GRAPHICAL:** characterised by a graphical representation encoding conditional dependencies and independencies
- **PROBABILISTIC:** described by a set of conditional probability distributions



$$X_1 \sim \mathcal{N}\left(\mu_{(1)} = \beta_0^{(1)}, \sigma^{2(1)}\right)$$

$$X_2 \mid X_1 \sim \mathcal{N}\left(\mu_{(2)} = \beta_0^{(2)} + \beta_1^{(2)} X_1, \sigma^{2(2)}\right)$$

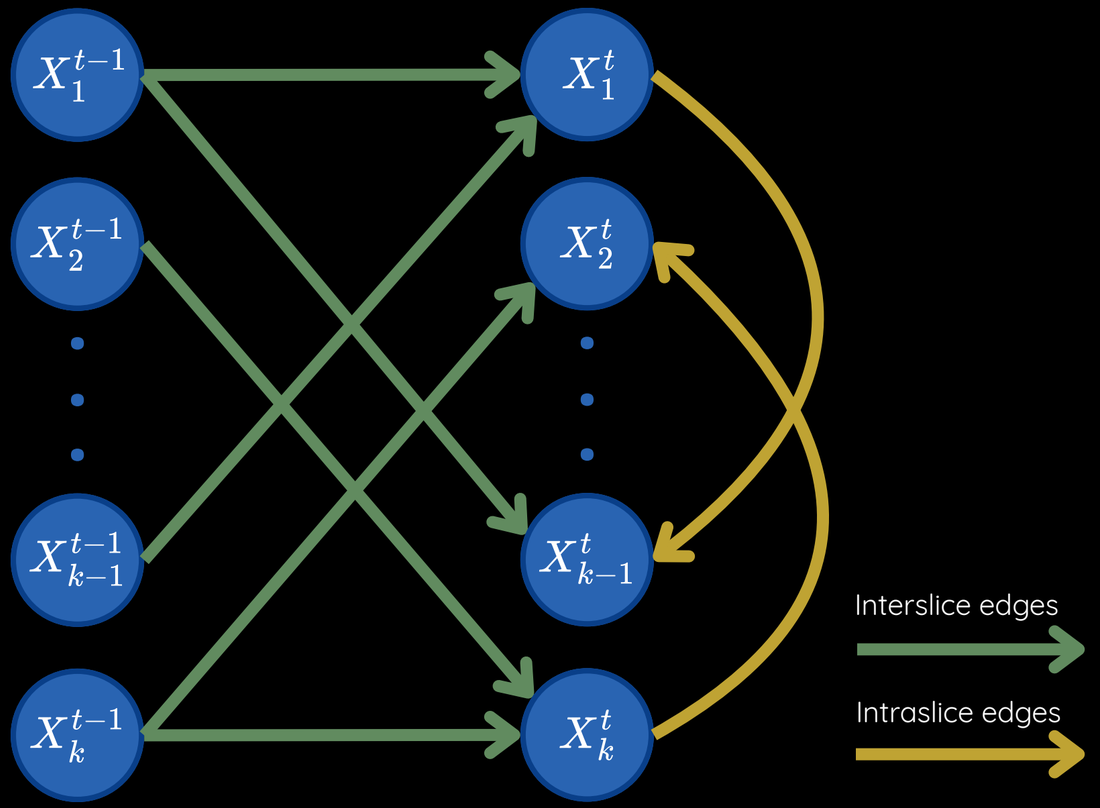
$$X_3 \mid X_1 \sim \mathcal{N}\left(\mu_{(3)} = \beta_0^{(3)} + \beta_1^{(3)} X_1, \sigma^{2(3)}\right)$$

$$X_4 \mid (X_1, X_3) \sim \mathcal{N}\left(\mu_{(4)} = \beta_0^{(4)} + \beta_1^{(4)} X_1 + \beta_2^{(4)} X_3, \sigma^{2(4)}\right)$$

Dynamic Bayesian Networks

- **Dynamic Bayesian networks** (DBNs) **model** the **temporal evolution** of **multivariate continuous time series** factorizing the joint distribution of all variables across time into a **product of conditional distributions for each node**

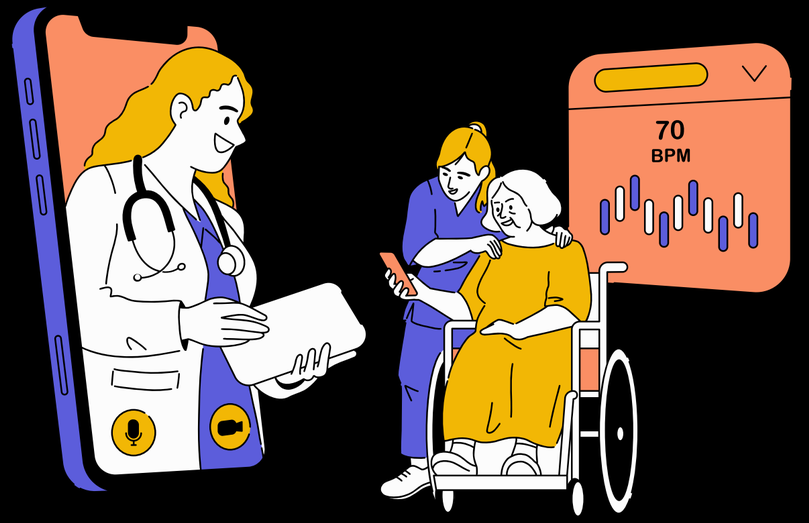
- **Dynamic Bayesian networks** (DBNs) are described as a set of:
  - A **structure**, represented by a **Transition Network** encoding the relationship between the **nodes at time t-1** and the **nodes at time t**
  - A **set of parameters** encoding the **linear relationships** between a variable and its parent set



Application domains and the context of Intensive Care Data

- DBNs are effective across **various domains** including **biology**, **healthcare domain** and **environmental data.** They have been used to learn gene regulatory networks, predict disease progression, and spatio-temporal environmental modeling.

- In the context of **Intensive Care Units** (ICUs) they show their value in **integrating heterogeneous ICU time-series data** and delivering **interpretable decision support** in the setting of **severe patients care.**



Bayesian Statistics
&
Markov Chain Monte Carlo

Frequentist Statistics vs Bayesian Statistics

Frequentist statistics *treats parameters as fixed but* ***unknown constants****. It evaluates the likelihood of observed data looking for the value of the* ***unknown parameters*** *which* ***maximize the likelihood****.*

$$\hat{\theta} = \max_{\theta} p(X|\theta)$$

Frequentist: “What parameter value best explains the observed data?”

Frequentist Statistics vs Bayesian Statistics

Frequentist statistics *treats parameters as fixed but* ***unknown constants****. It evaluates the likelihood of observed data looking for the value of the* ***unknown parameters*** *which* ***maximize the likelihood****.*

$$\hat{\theta} = \max_{\theta} p(X|\theta)$$

<u>Frequentist</u>: “What parameter value best explains the observed data?”

Bayesian statistics *views* ***parameters as random variables with probability distributions****.* ***Prior beliefs*** *about these parameters are* ***updated with observed data*** *to produce* ***posterior distributions****. It directly* ***model uncertainty over the parameters*** *through their distributions.*

$$p(\theta|X) = \frac{p(\theta) \cdot p(X|\theta)}{p(X)}$$

<u>Bayesian</u>: “Given the data, what is our updated belief over the parameters probability distribution?”

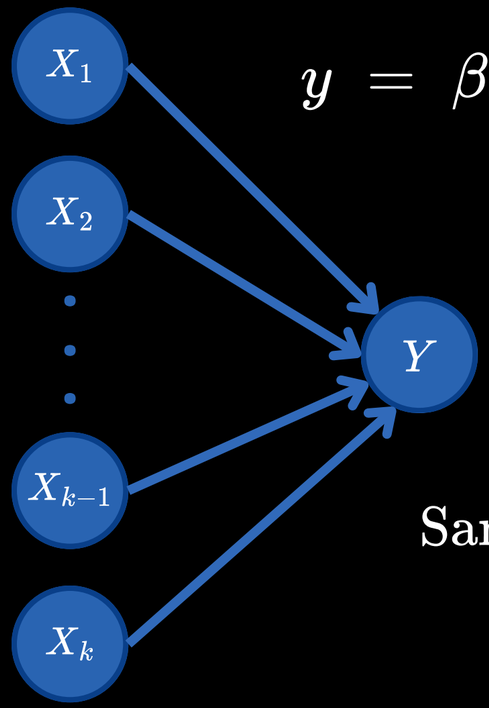
Markov Chain Monte Carlo (MCMC) Simulations in Bayesian Statistics

Bayesian inference strongly relies on MCMC simulations to approximate posterior expectations when the required integrals cannot be evaluated in closed form.

Markov Chain Monte Carlo (MCMC) Simulations in Bayesian Statistics

Standard MCMC Setting

Conditional to the **known model** (structure) the **unknown parameters** are iteratively sampled



$$y = \beta_0 + \sum_{j:1:k} \beta_j x_j + \epsilon_i$$

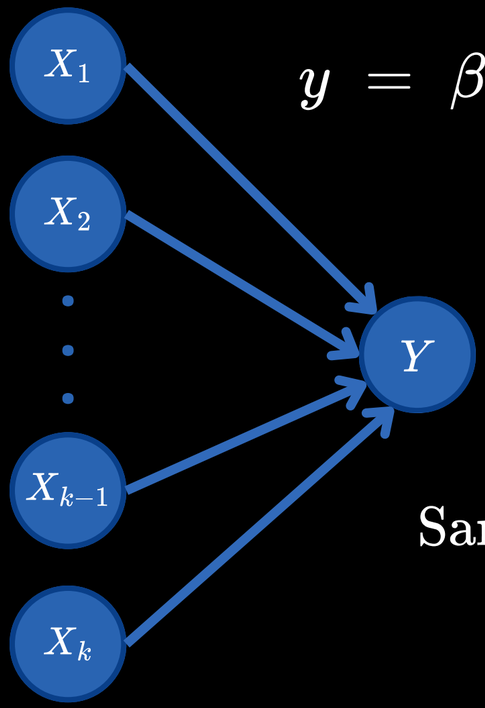
$$\text{Sample } \beta = \begin{pmatrix} \beta_0 \\ \vdots \\ \beta_k \end{pmatrix}$$

Bayesian inference strongly relies on MCMC simulations to approximate posterior expectations when the required integrals cannot be evaluated in closed form.

Markov Chain Monte Carlo (MCMC) Simulations in Bayesian Statistics

Standard MCMC Setting

Conditional to the **known model** (structure) the **unknown parameters** are iteratively sampled

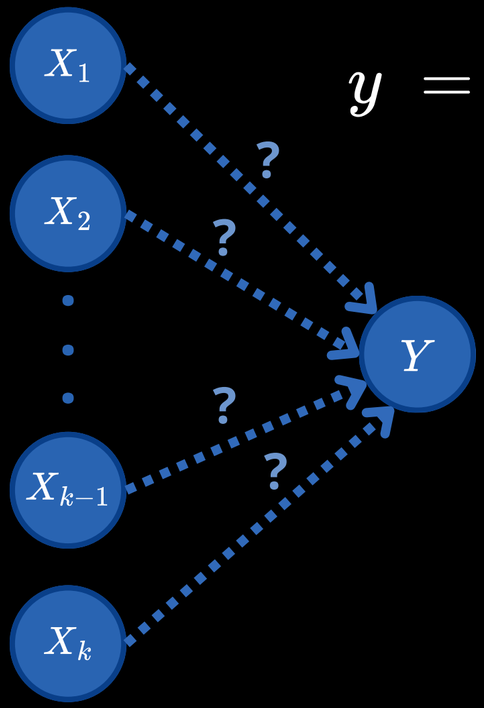


$$y = \beta_0 + \sum_{j:1:k} \beta_j x_j + \epsilon_i$$

$$\text{Sample } \beta = \begin{pmatrix} \beta_0 \\ \vdots \\ \beta_k \end{pmatrix}$$

RJMCMC Setting

The **unknown parameters** are sampled jumping between different **unknown models**



$$y = \beta_0 + \sum_{j:(X_j \in \pi)} \beta_j x_j + \epsilon_i$$

$$\text{Move to a model } \pi$$

$$\text{Sample } \beta = [\beta_j]_{j:(X_j \in k)}$$

Bayesian inference strongly relies on MCMC simulations to approximate posterior expectations when the required integrals cannot be evaluated in closed form.

Metropolis-Hastings & Gibbs Sampling

Metropolis–Hastings (MH) is an MCMC algorithm that constructs a Markov chain by **proposing candidate values**, **sampled from a proposal distribution** (PD).

1. Initialization of the chain: θ_0
2. Proposal of a new value sampled from the PD: $\theta_\star \sim q(\theta_\star|\theta_{t-1})$
3. Compute the Acceptance probability: $\alpha = \min\left(1, \frac{p(\theta_\star \mid \cdot)\, q(\theta_{t-1} \mid \theta^*)}{p(\theta_{t-1} \mid \cdot)\, q(\theta_\star \mid \theta_{t-1})}\right)$
4. Accept the proposed value with $\alpha > p \sim U(0,1)$ otherwise keep the previous value θ_{t-1}

Metropolis-Hastings & Gibbs Sampling

Metropolis–Hastings (MH) is an MCMC algorithm that constructs a Markov chain by **proposing candidate values**, **sampled from a proposal distribution** (PD).

1. Initialization of the chain: θ_0
2. Proposal of a new value sampled from the PD: $\theta_\star \sim q(\theta_\star|\theta_{t-1})$
3. Compute the Acceptance probability: $\alpha = \min\left(1, \frac{p(\theta_\star \mid \cdot)\, q(\theta_{t-1} \mid \theta^*)}{p(\theta_{t-1} \mid \cdot)\, q(\theta_\star \mid \theta_{t-1})}\right)$
4. Accept the proposed value with $\alpha > p \sim U(0,1)$ otherwise keep the previous value θ_{t-1}

Gibbs sampling (GS) is a **special case of Metropolis–Hastings** where the **PD is the full conditional distribution** (FCD), namely the distribution of the parameter given the current model and data.

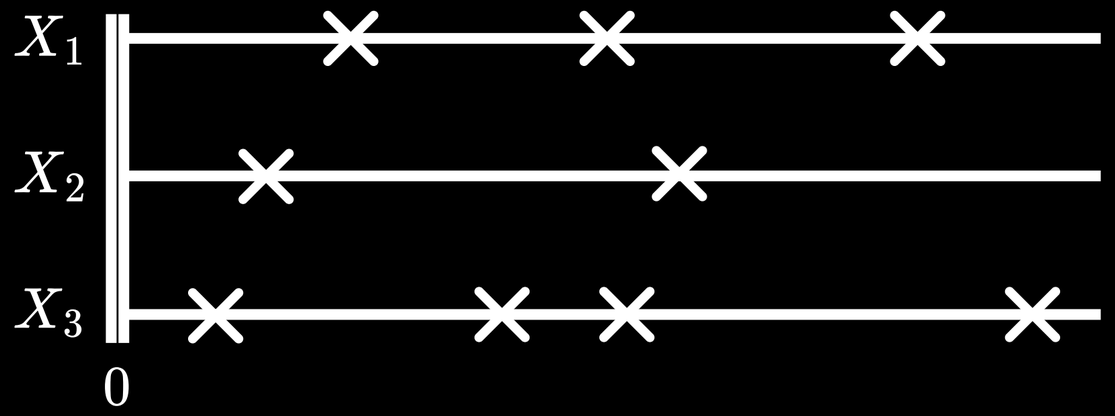
In GS the acceptance probability $\alpha = \min\left(1, \frac{p(\theta_\star \mid \cdot)\, q(\theta_{t-1} \mid \theta^*)}{p(\theta_{t-1} \mid \cdot)\, q(\theta_\star \mid \theta_{t-1})} = \frac{p(\theta_\star \mid \cdot)\, p(\theta_{t-1} \mid \cdot)}{p(\theta_{t-1} \mid \cdot)\, p(\theta_\star \mid \cdot)}\right) = 1$

However **sampling exactly from the FCD is generally intractable** except for conjugate priors frameworks.

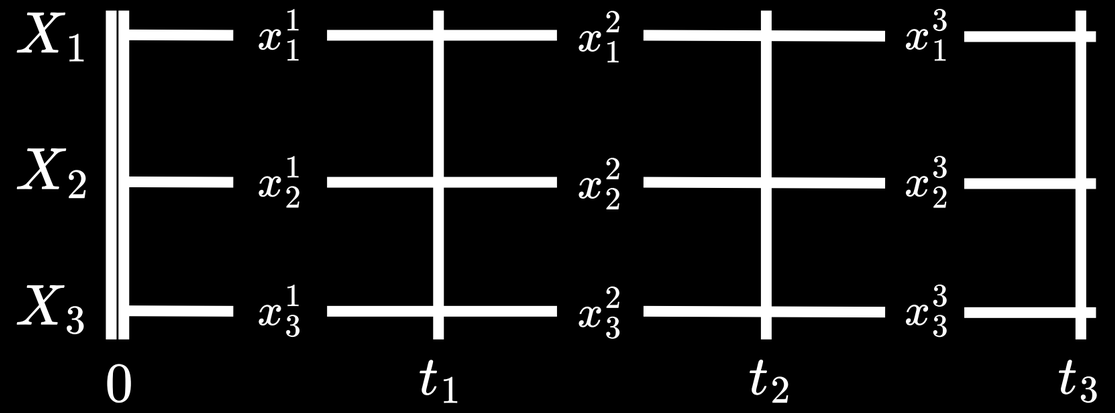
Issues in Observational Settings

DBNs in Observational Settings

OBSERVATIONAL DATA SCENARIO

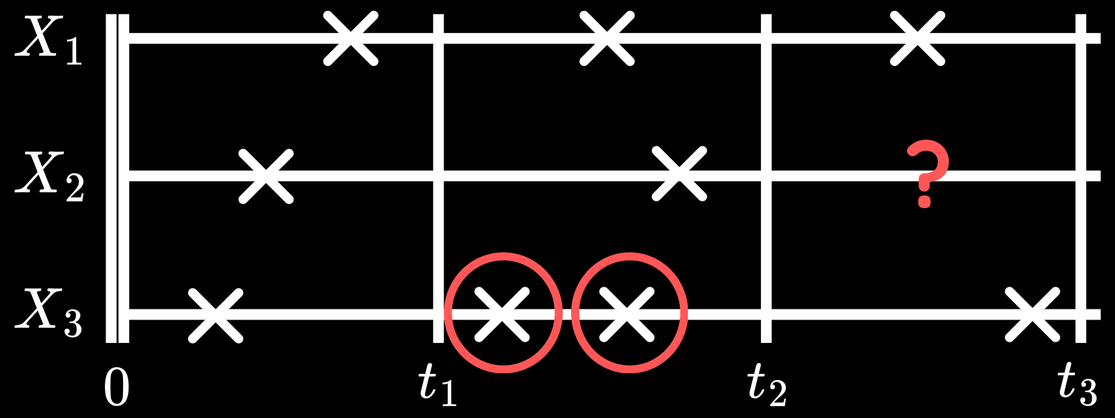


INPUT DATA REQUIRED IN DBNs



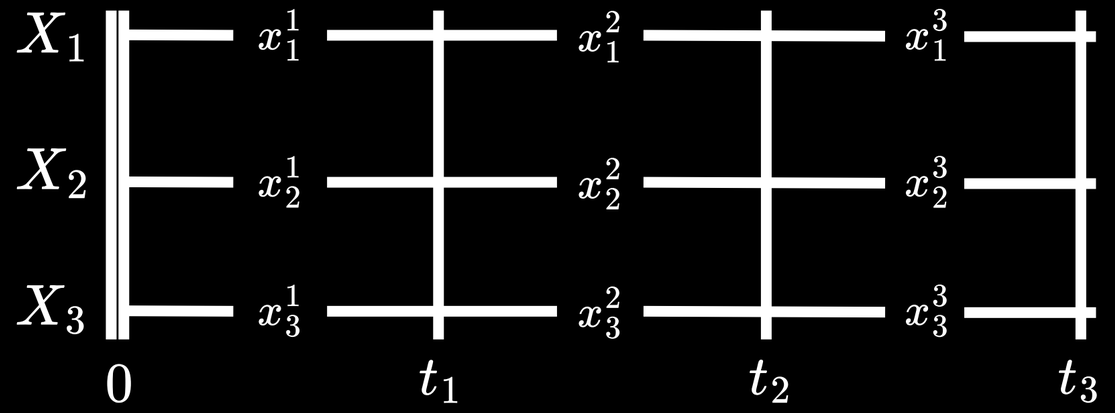
DBNs in Observational Settings

OBSERVATIONAL DATA SCENARIO

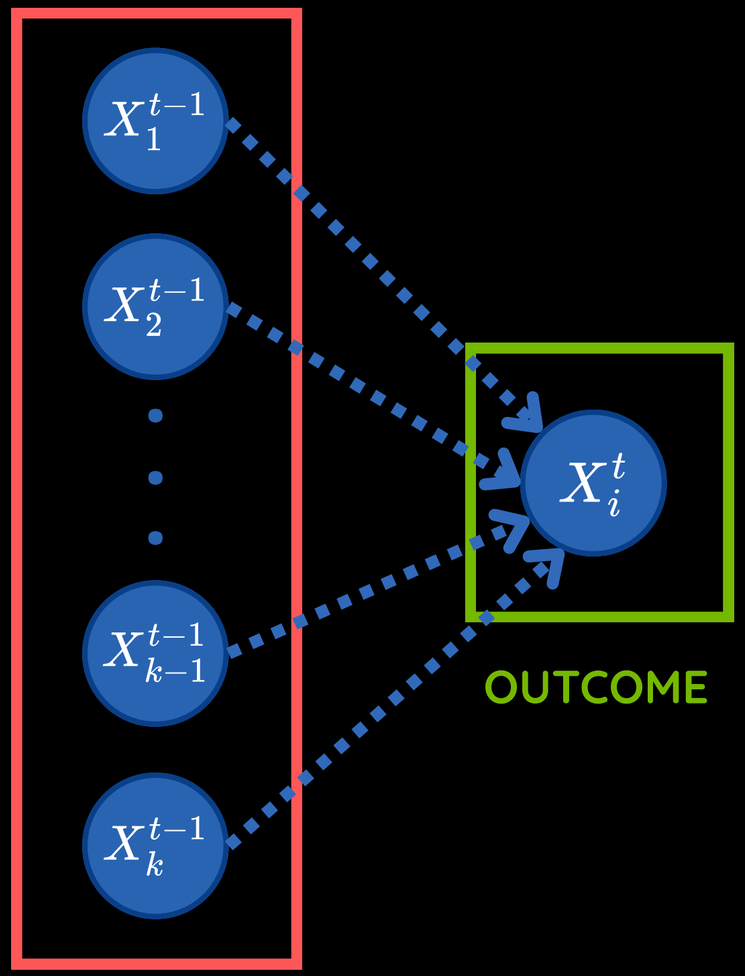


<< How can we manage observations over continuous time in DBN learning? >>

INPUT DATA REQUIRED IN DBNs



Dynamic Bayesian Networks as a set of Linear Regressions



$$x_i^t = \beta_0^{(i)} + \sum_{j:(X_j^{t-1} \in \pi_{(i)})} \beta_j^{(i)} x_j^{t-1} + \epsilon_i^t$$

$$(t = 1, \ldots, T)$$

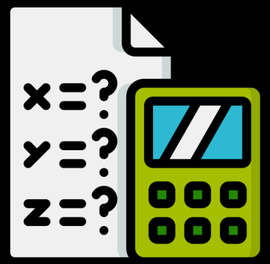
NO ACYCLICITY CONSTRAINT IS NEEDED AS IN STATIC BNS !

Managing Missing Values in DBNs

ICU data are characterized by **irregular sampling frequencies** and incorrect detection, leading to **Incomplete Data**

STANDARD IMPUTATION

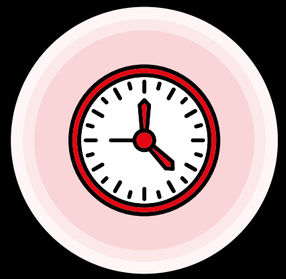
i.e. MICE Imputation



- ⊕ **REALLY FAST**
- ⊖ **MODEL AGNOSTIC**
- ⊖ **POINT ESTIMATES**

FREQUENTIST APPROACHES

i.e. Structural EM



- ⊕ **ITERATIVE UPDATES**
- ⊕ **MODEL-DEPENDENT**
- ⊖ **LOCAL MINIMA RISK**
- ⊖ **NOT DBN-SPECIFIC**

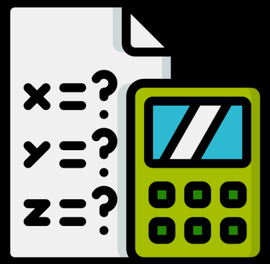
STATE OF THE ART

Managing Missing Values in DBNs

ICU data are characterized by **irregular sampling frequencies** and incorrect detection, leading to **Incomplete Data**

STANDARD IMPUTATION

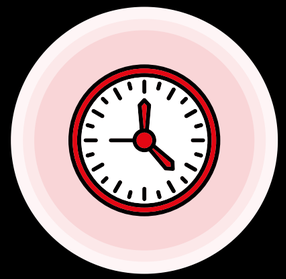
i.e. MICE Imputation



- ⊕ **REALLY FAST**
- ⊖ **MODEL AGNOSTIC**
- ⊖ **POINT ESTIMATES**

FREQUENTIST APPROACHES

i.e. Structural EM



- ⊕ **ITERATIVE UPDATES**
- ⊕ **MODEL-DEPENDENT**
- ⊖ **LOCAL MINIMA RISK**
- ⊖ **NOT DBN-SPECIFIC**

STATE OF THE ART

FULL BAYESIAN METHOD

LUME-DBN

- ⊕ **MODEL UNCERTAINTY**
- ⊕ **AVOID LOCAL MINIMA**
- ⊕ **DBN-SPECIFIC**
- ? **CONVERGENCE TO THE RIGHT SOLUTION**

OUR PROPOSAL

Learning DBNs from Incomplete Data

A Bayesian Formulation of Dynamic Bayesian Networks

Since intra-slice edges are ruled out, the task of learning a DBN could be segmented in k distinct BLR learning tasks

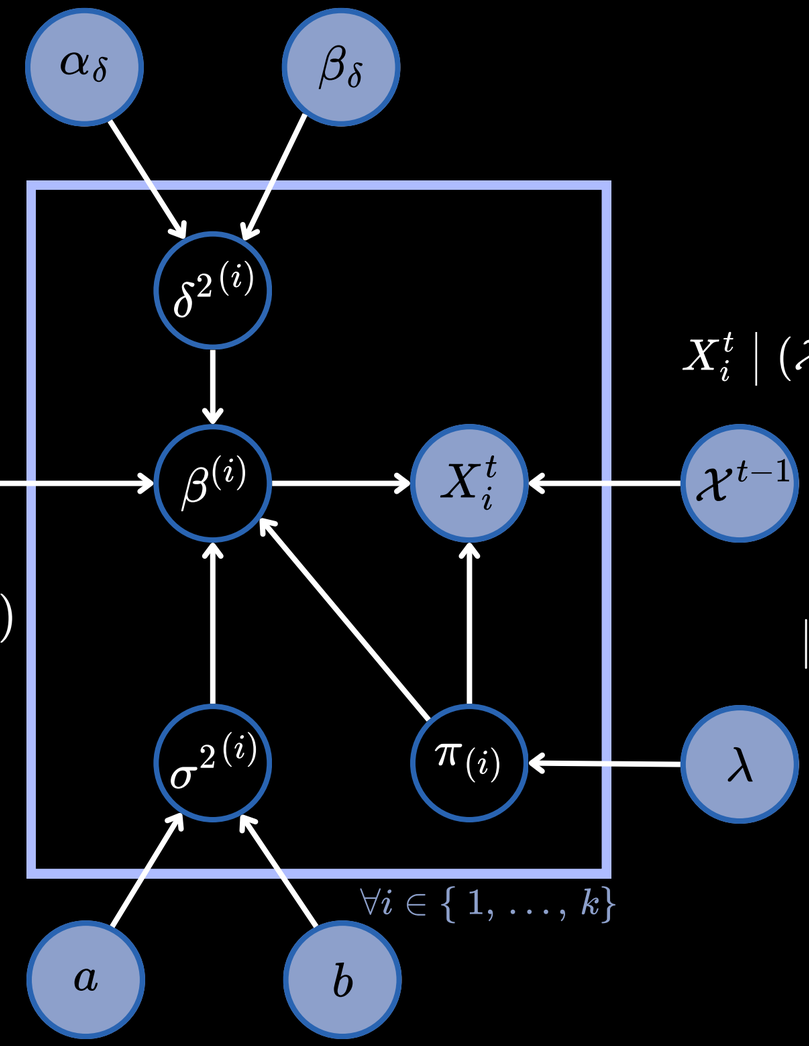
$$\delta^{2(i)} \sim \text{Inv-Gamma}(\alpha_\delta, \beta_\delta)$$

$$X_i^t \mid (\mathcal{X}^{t-1}, \beta^{(i)}, \sigma^{2(i)}, \pi_{(i)}) \sim \mathcal{N}\left(\mathcal{X}^{t-1}_{[\pi_{(i)}]}\, \beta^{(i)}, \sigma^{2(i)} I\right)$$

$$\beta^{(i)} | (\sigma^{2(i)}, \delta^{2(i)}, \pi_{(i)}) \sim \mathcal{N}(\mu_{[\pi_{(i)}]}, \sigma^{2(i)} \delta^{2(i)} I)$$

$$|\pi_{(i)}| \sim \text{Poisson}(\lambda)$$

$$\sigma^{2(i)} \sim \text{Inv-Gamma}(a, b)$$

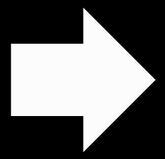


Learning DBNs from Incomplete Data

Learning the DBN is then the task of **learning** the **structure** and the set of **parameters** that describe the **relationships between temporal nodes** in the Transition Network. To **learn** a **DBN from data**, we consider a group of **N samples partially observed** at **T+1 equally spaced time points**, where each sample is described by a **set of k variables.**

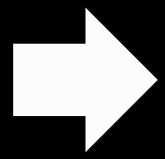
INITIAL DATASET

Sample	Time	**X1**	**X2**
s1	0.3	?	3
s1	1.2	2	5
s1	1.8	4	7
s2	0.9	9	?
s2	1.5	4	1



TEMPORALLY DISCRETIZED DATA

Sample	Time	**X1**	**X2**
s1	0	2	3
s1	1	3	6
s2	0	9	?
s2	1	1	4



LAGGED DATA

Sample	Time	**X1 (t-1)**	**X1 (t)**	**X2 (t-1)**	**X2 (t)**
s1	1	2	3	3	6
s2	1	9	1	?	4

LUME-DBN Algorithm

Latent Uncertainty Modeling via MCMC Estimation in DBNs (**LUME-DBN**) is the first **full bayesian method** for **learning DBNs** from **Incomplete data.** Employing our Gibbs sampling imputation step, LUME-DBN is characterized by an iterative procedure following **three main steps**:

For each outcome variable

PARAMETERS UPDATES: COLLAPSED GIBBS SAMPLING STEP

STRUCTURAL UPDATES: METROPOLIS-HASTING STEP

MISSING VALUES IMPUTATION: GIBBS SAMPLING STEP

For each sample and time frame

Structure and Parameter Updates

Conditional on the current structure and missing values

Sampling $\sigma^2, \beta, \delta^2$ from their Full Conditional Distributions (FCDs)

$$
\begin{cases}
\sigma^2 \sim \mathrm{Inv-GAM}\Big(\alpha_\sigma + \dfrac{NT}{2},\ \beta_\sigma + \dfrac{1}{2}(Y - X_{[\pi]}\mu_{[\pi]})^\top (I + \sigma^2 X_{[\pi]} X_{[\pi]}^\top)^{-1}(Y - X_{[\pi]}\mu_{[\pi]})\Big) \\
\beta \sim \mathcal{N}\Big((\sigma^{-2} I + X_{[\pi]}^\top X_{[\pi]})^{-1}(\sigma^{-2}\mu_{[\pi]} + X_{[\pi]}^\top Y),\ \sigma^2(\sigma^{-2} I + X_{[\pi]}^\top X_{[\pi]})^{-1}\Big), \\
\delta^2 \sim \mathrm{Inv-GAM}\Big(a + \dfrac{|\pi| + 1}{2},\ b + \dfrac{1}{2}\sigma^{-2}(\beta - \mu_{[\pi]})^\top (\beta - \mu_{[\pi]})\Big)
\end{cases}
$$

Structure and Parameter Updates

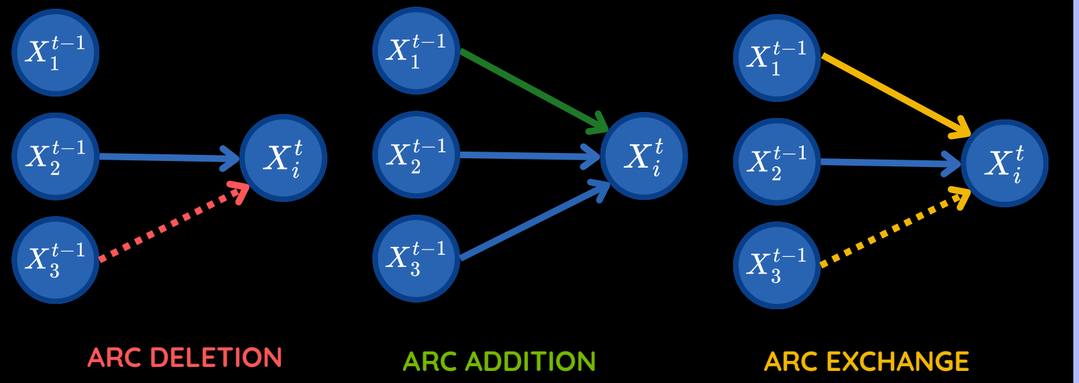
Conditional on the current structure and missing values

Sampling $\sigma^2, \beta, \delta^2$ from their Full Conditional Distributions (FCDs)

$$\begin{cases} \sigma^2 \sim \text{Inv}-\text{GAM}\Big(\alpha_\sigma + \dfrac{NT}{2},\ \beta_\sigma + \dfrac{1}{2}(Y - X_{[\pi]}\mu_{[\pi]})^\top (I + \sigma^2 X_{[\pi]} X_{[\pi]}^\top)^{-1}(Y - X_{[\pi]}\mu_{[\pi]})\Big) \\ \beta \sim \mathcal{N}\Big((\sigma^{-2} I + X_{[\pi]}^\top X_{[\pi]})^{-1}(\sigma^{-2}\mu_{[\pi]} + X_{[\pi]}^\top Y),\ \sigma^2(\sigma^{-2} I + X_{[\pi]}^\top X_{[\pi]})^{-1}\Big), \\ \delta^2 \sim \text{Inv}-\text{GAM}\Big(a + \dfrac{|\pi| + 1}{2},\ b + \dfrac{1}{2}\sigma^{-2}(\beta - \mu_{[\pi]})^\top (\beta - \mu_{[\pi]})\Big) \end{cases}$$

Conditional on the imputed missing values

Sampling a Move and Propose a new Structure



Structural Update Probability of Acceptance

$$A(\pi \to \pi_\star) = \min\left\{1,\ \frac{p(Y \mid \pi_\star, \delta^2)}{p(Y \mid \pi, \delta^2)} \cdot \frac{p(\pi_\star)}{p(\pi)} \cdot HR\right\}$$

$$\text{Acceptance for } A > p \sim U(0,1)$$

Missing Values Update - Gibbs Sampling Step

Conditional on the current structure and parameters

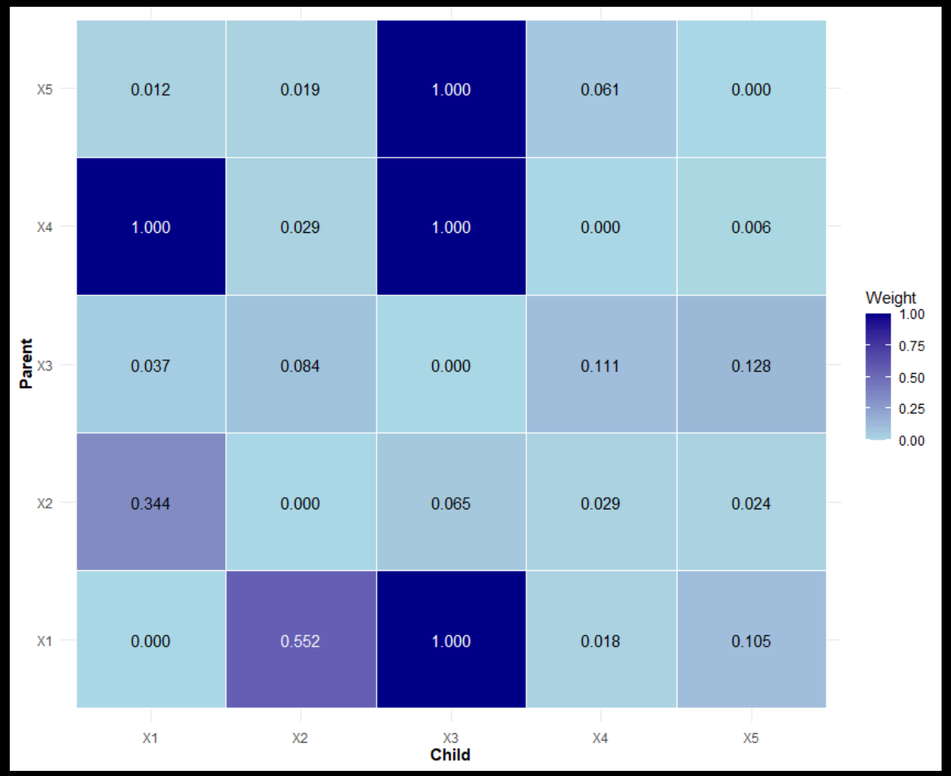
Compute Missing Values Posterior Parameters

$$\begin{cases} \mu^* = \sigma^{2*} \cdot \left(\frac{\mu_i}{\sigma^{2(i)}} + \sum_{j:(X_i^t \in \pi_{(j)})} \beta_i^{(j)} \cdot \frac{x_j^{t+1} - \mu_{\{-i\}}^{(j)}}{\sigma^{2(j)}} \right) \\ \sigma^{2*} = \left(\frac{1}{\sigma^{2(i)}} + \sum_{j:(X_i^t \in \pi_{(j)})} \frac{\left(\beta_i^{(j)}\right)^2}{\sigma^{2(j)}} \right)^{-1} \end{cases}$$

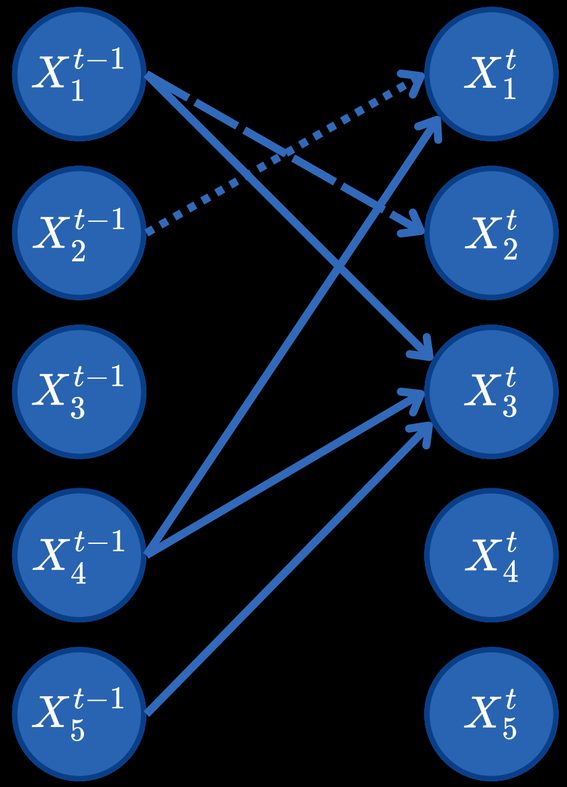
Sampling Missing Values iteratively from their FCD

$$x_{i\,[MIS]} \sim \mathcal{N}(\mu^*, \sigma^{2*})$$

Posterior Distributions



Edge Indicators Posteriors



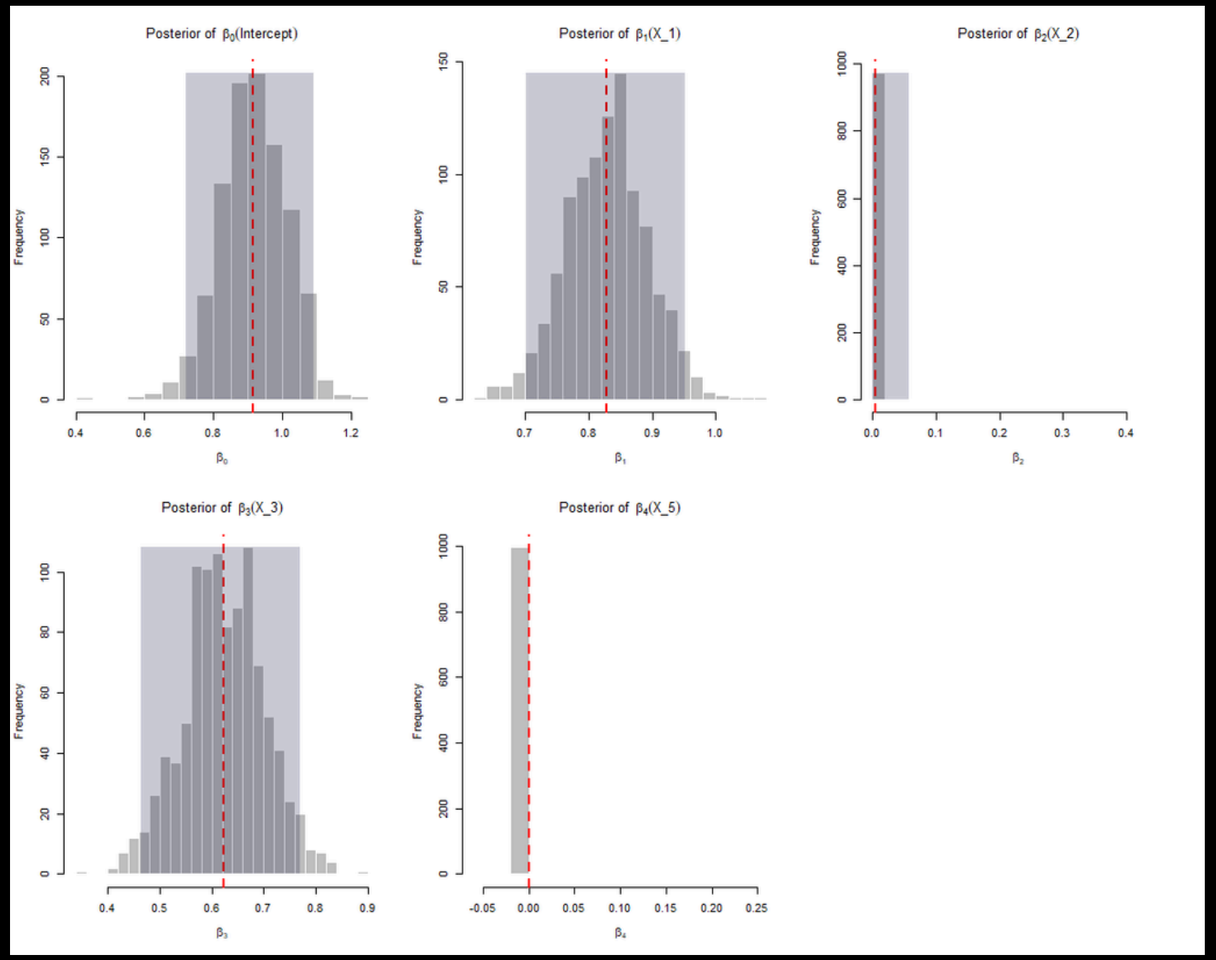
DBN Structure

FUN IN RESTRICTION → KEEP ONLY THE LAST M SAMPLES (AFTER CONVERGENCE)

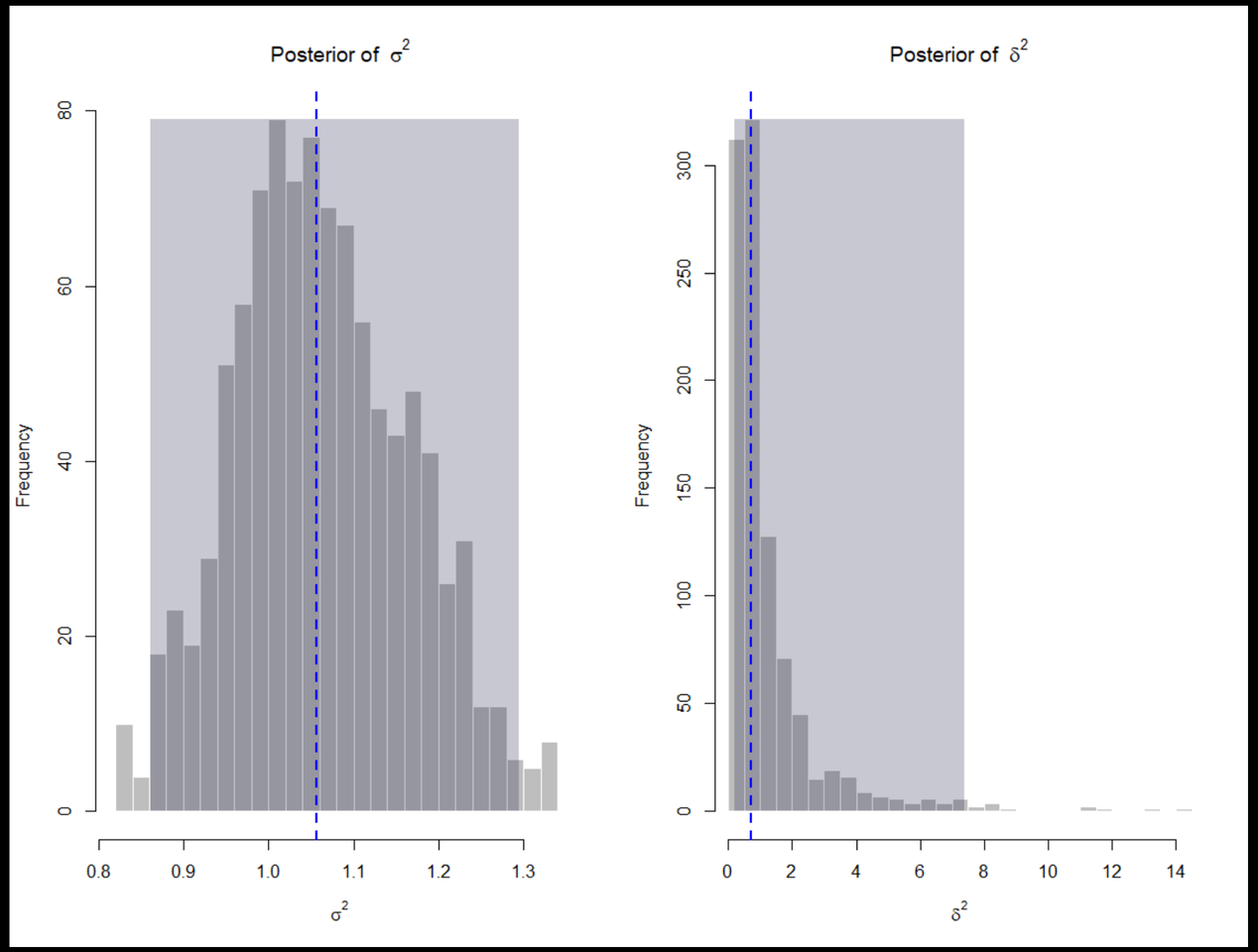
THIN OUT STEP → KEEP ONLY ONE SAMPLE EVERY υ TO REDUCE AUTOCORRELATION

Posterior Distributions

For each Bayesian Linear Regression:



Linear Coefficients Posteriors



Noise and Uncertainty Parameters Posteriors

Applications in Intensive Care

LUME-DBN Validation

SIMULATION STUDIES

ICU CHALLENGE DATA

PREMATURITY CLINICAL DATA

LUME-DBN Validation

SIMULATION STUDIES

+ DBN REFERENCE STRUCTURE AVAILABLE

+ DIFFERENT EXPERIMENTAL SCENARIOS

– DATA IS UNREALISTIC

ICU CHALLENGE DATA

PREMATURITY CLINICAL DATA

Simulated Experiments - Experimental Setting

To assess the validity of LUME-DBN both in terms of convergence and network reconstruction accuracy multiple incomplete datasets are generated from 10 distinct DBNs

DBN GENERATION

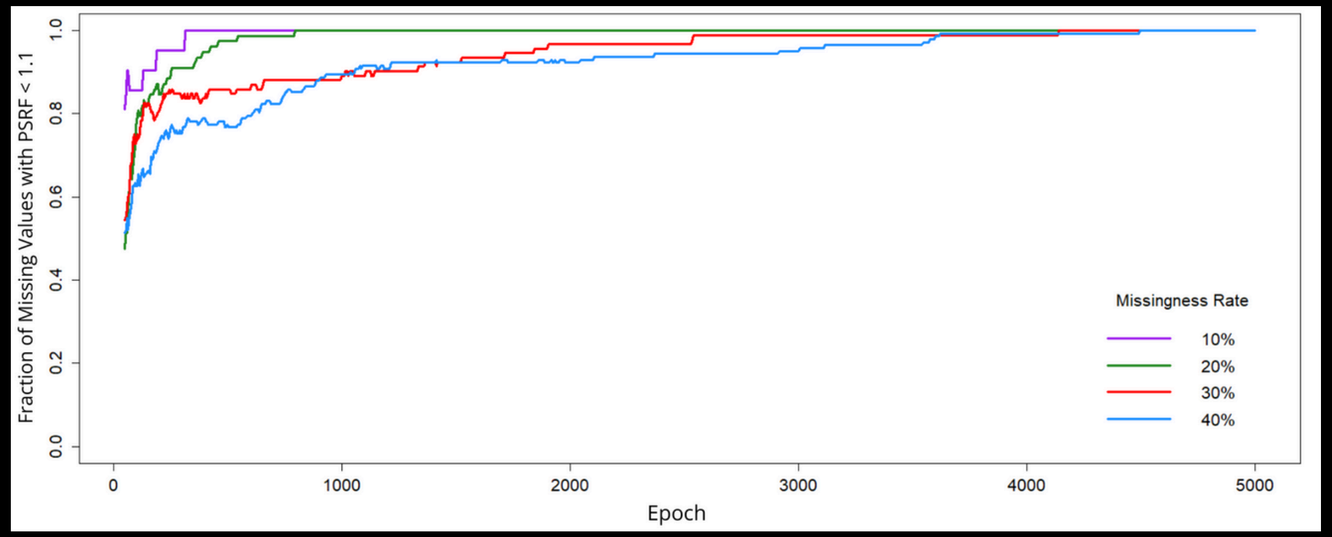
- **10 DBNs** with **10 nodes** and a maximum of 5 parents per node
- Random Topological Order to avoid
- Parameters: $\beta \in U(0.2, 0.5) - \sigma^2 = 1$

DATA GENERATION

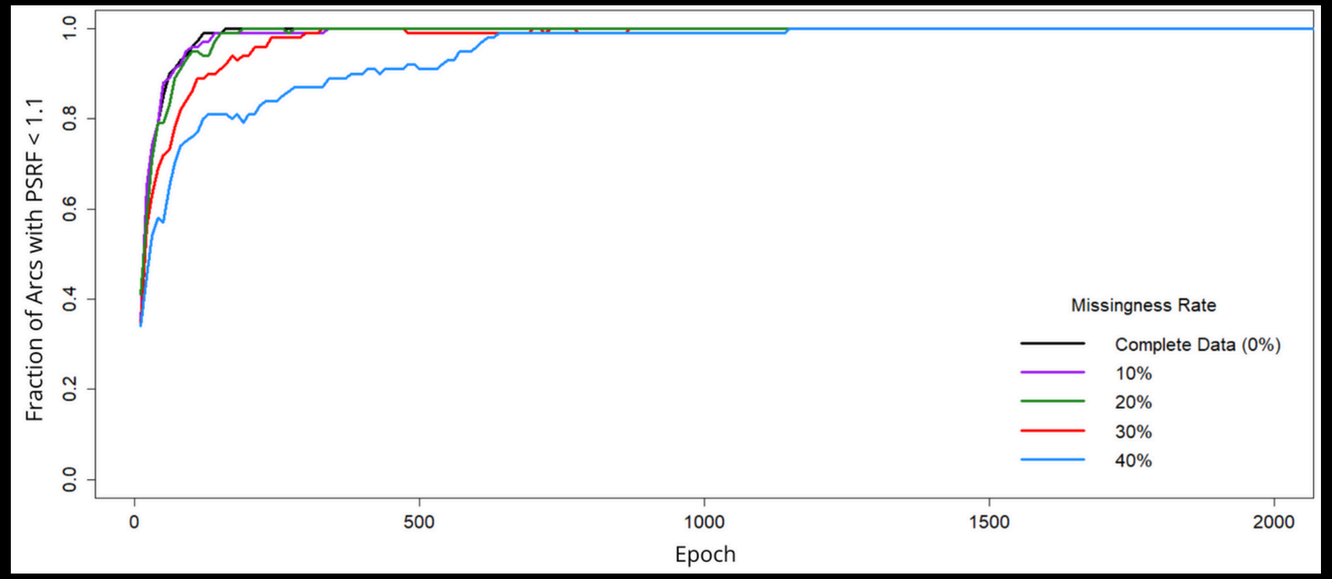
- For each DBN a set of random time series is generated with 3 sample sizes: $T = \{50, 100, 200\}$
- For each complete Dataset 4 incomplete dataset are generated at 4 sample sizes: $\{0.1, 0.2, 0.3, 0.4\}$

Simulation Results - Convergence Diagnostics

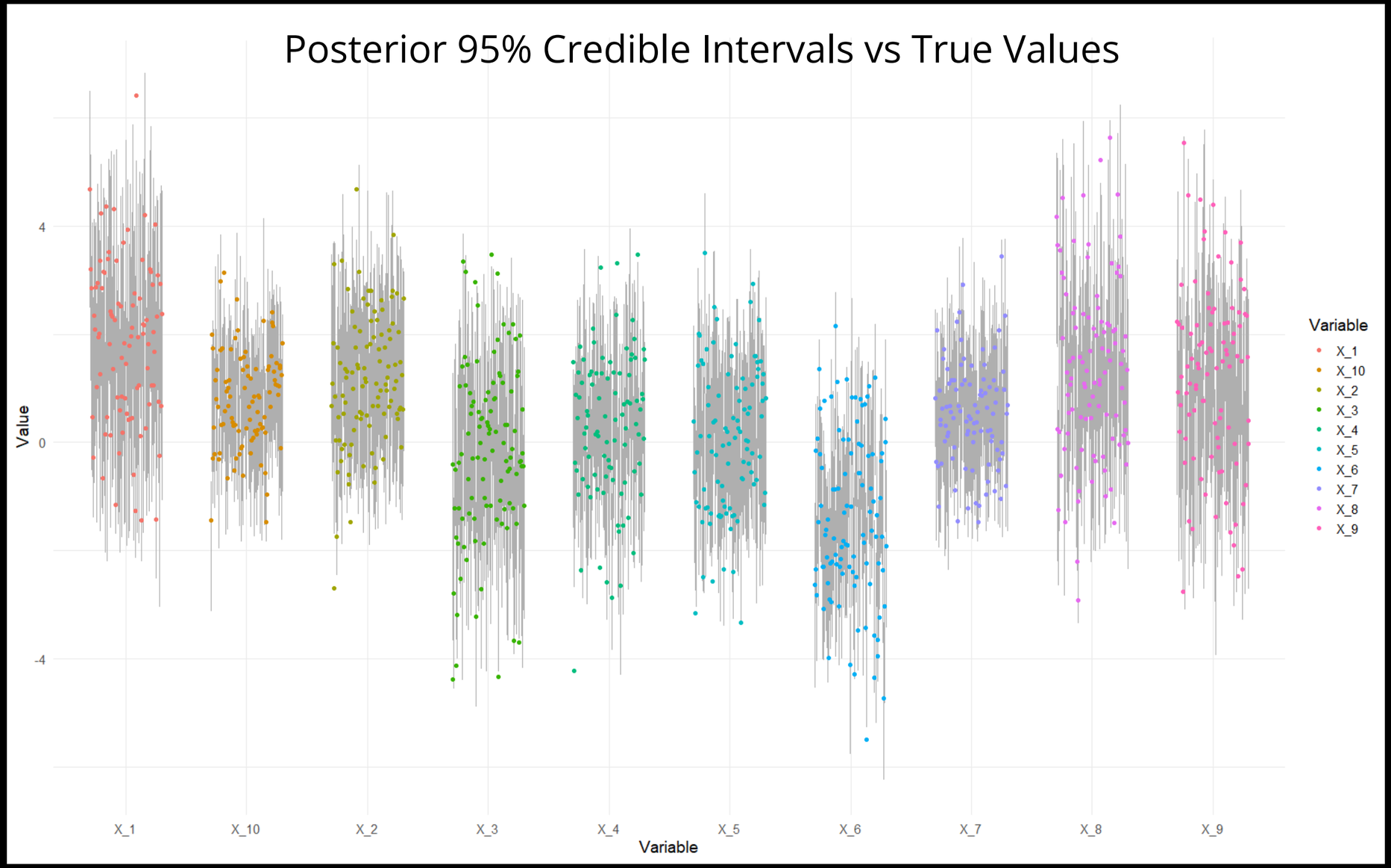
The convergence of the **DBN structure** and the **missing values is evaluated in terms of potential scale reduction factors** along parallel simulations



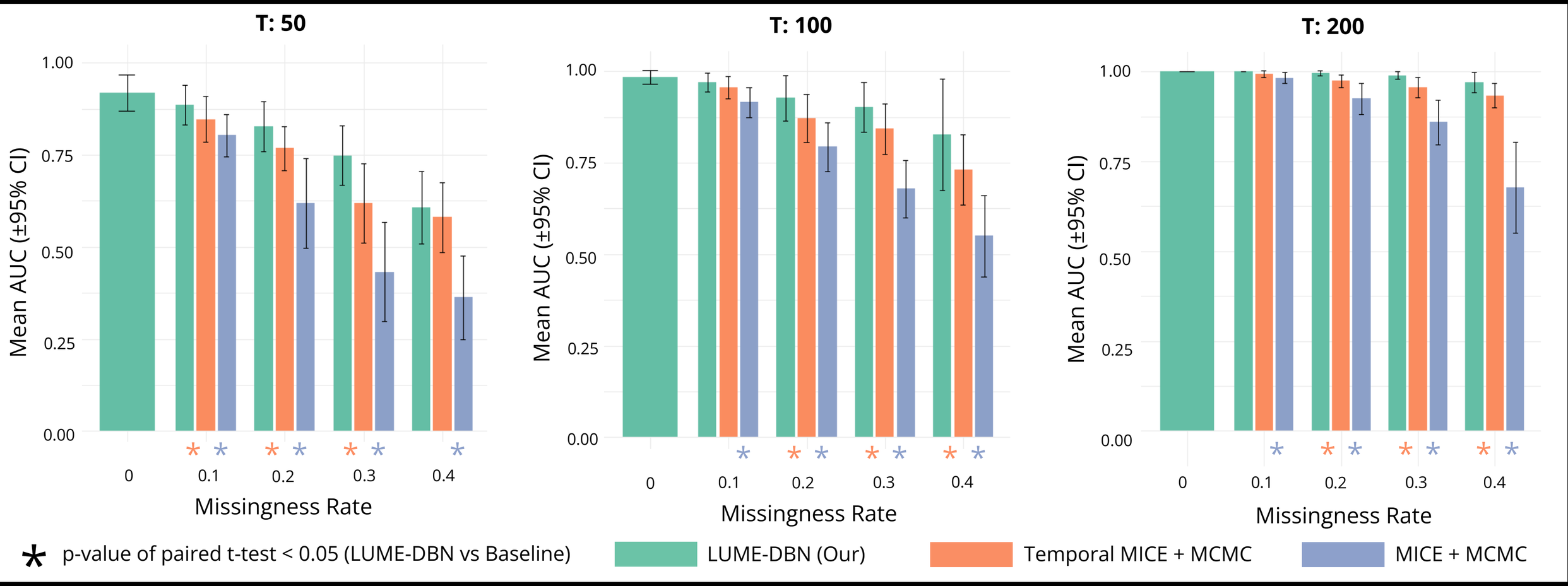
Convergence is reached when the **proportion** of **posterior samples** with **PSRF < 1.1** consistently **reaches 1**



Simulation Results - 95% Credible Intervals for Missing Values



Simulation Results - Network Reconstruction Accuracy



LUME-DBN Validation

SIMULATION STUDIES

ICU CHALLENGE DATA

PREMATURITY CLINICAL DATA

+ REAL-WORLD RETROSPECTIVE DATA

+ DATA FROM MULTIPLE FACILITIES

– UNKNOWN DATA COLLECTION PROCEDURE

– NO REFERENCE STRUCTURE

A case study in Critical Care - The Physionet 2012 Dataset

PHYSIONET DATABASE

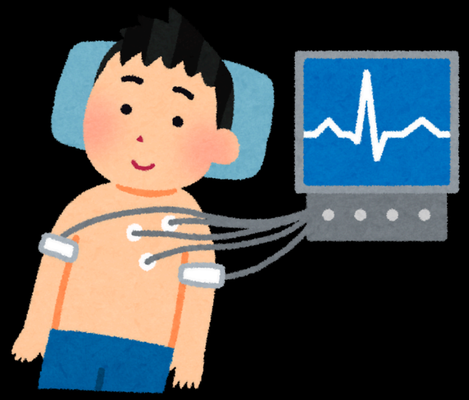
- Records from 20k patients admitted to the ICU
- Data acquired in the first 48 hours of ICU stay

SAMPLE SELECTION

- Including severe patients only (SOFA score > 12)
- Stratification based on ICU types

MEDICAL ICU (MICU)
34 patients

SURGICAL ICU (SICU)
104 patients



CORONARY CARE (CCU)
114 patients

CARDIAC SURGERY (CSRU)
62 patients

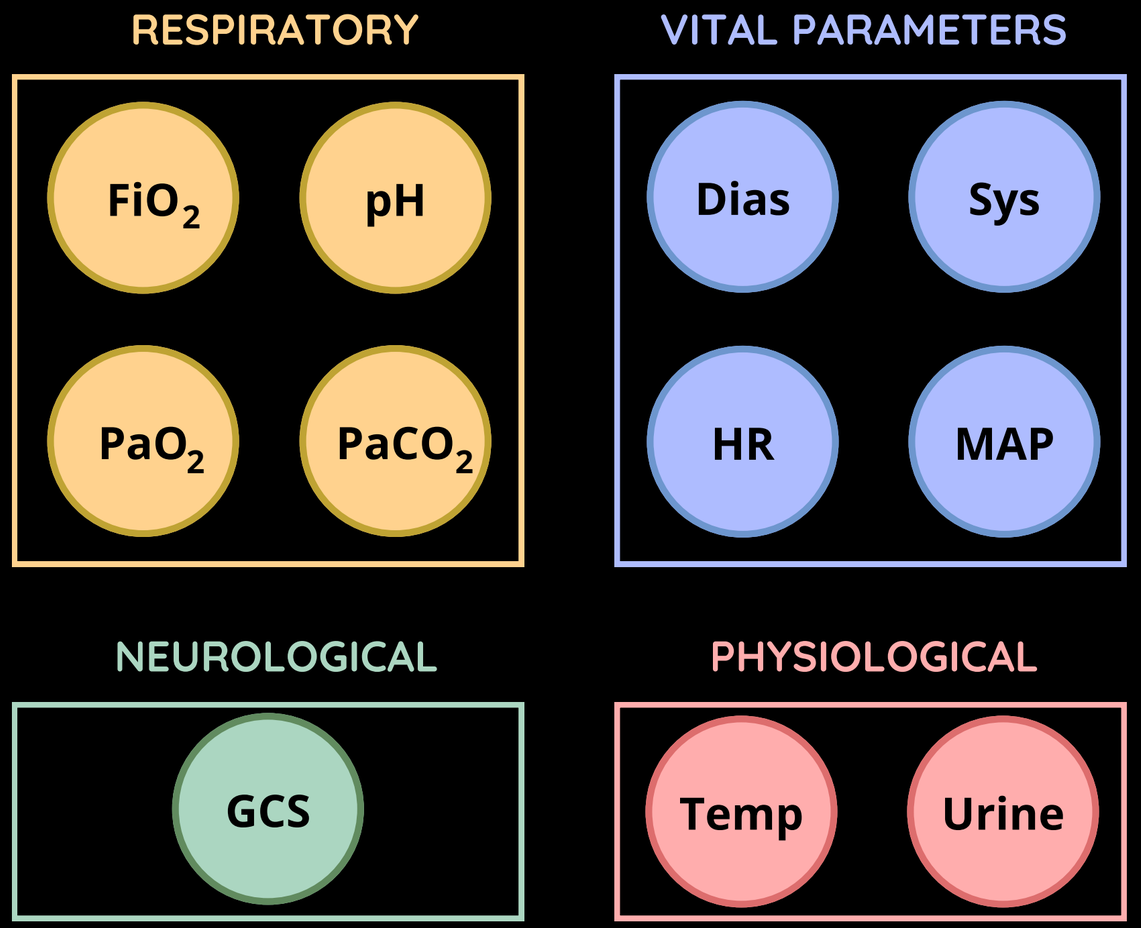
A case study in Critical Care - Different Standardization Techniques

TEMPORAL DISCRETIZATION

- Discretization into 6-hour intervals (T = 9)
- Keeping the variables with missingness rates < 40% compatible with the experimental results

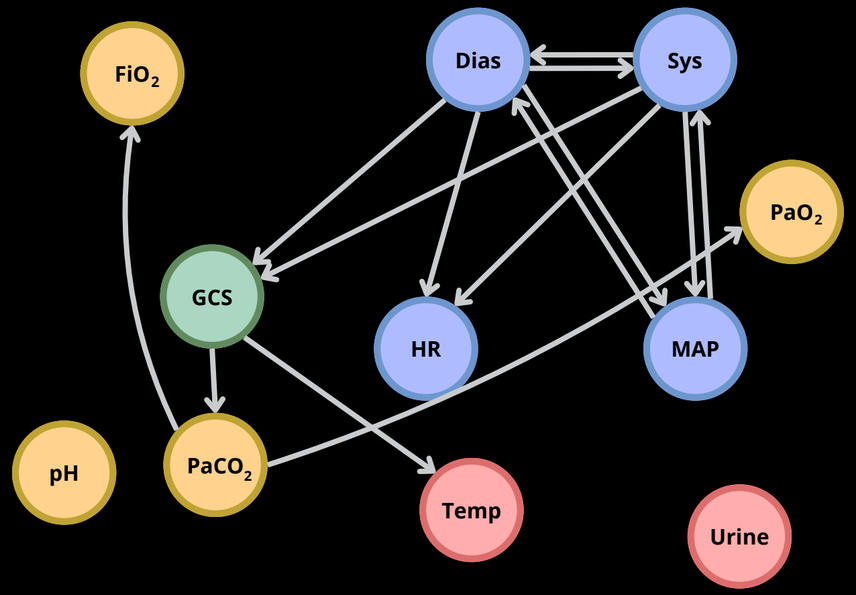
STANDARDIZATION

- Local Standardization: applied for each ICU group
- Global Standardization: on the whole dataset

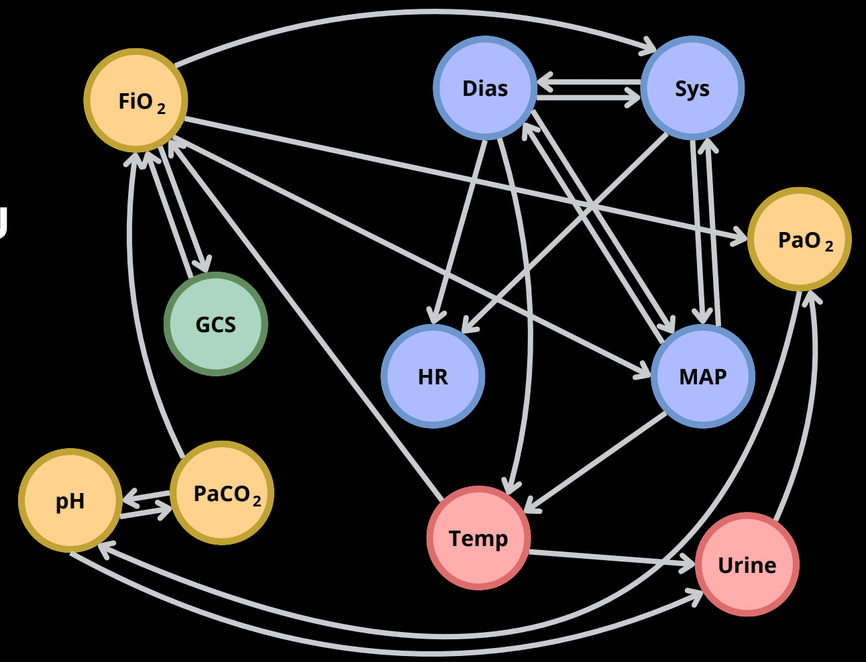


Commonalities and Differences between group-specific DBNs

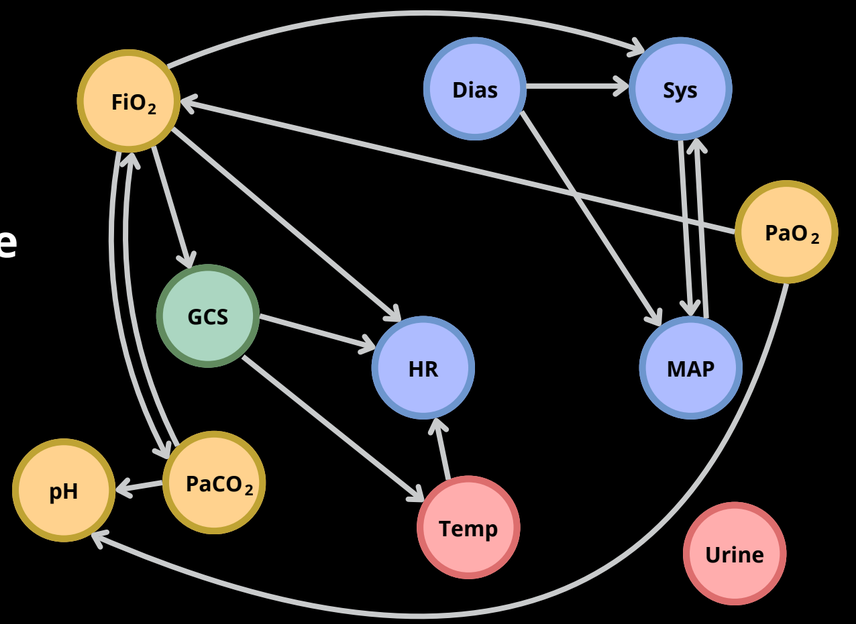
Medical ICU (MICU)



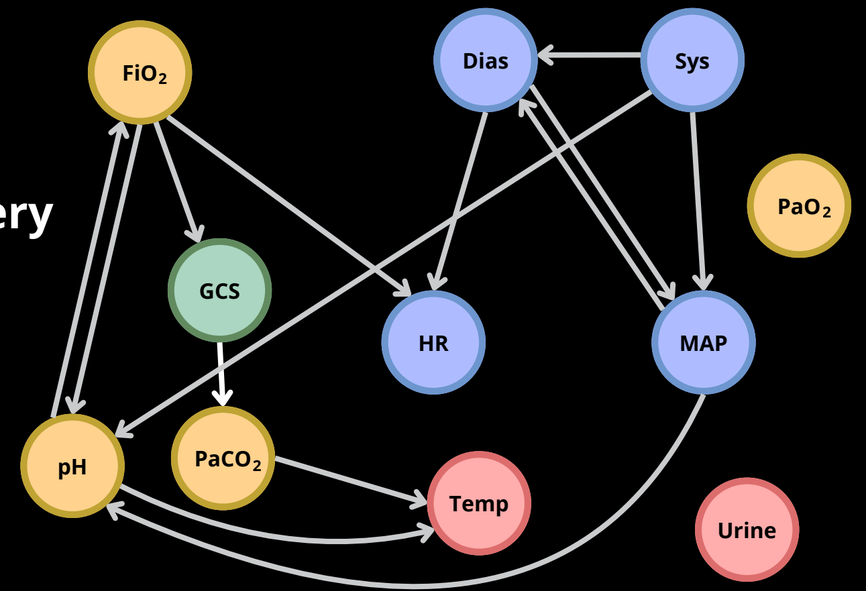
Surgical ICU (SICU)



Coronary Care (CCU)

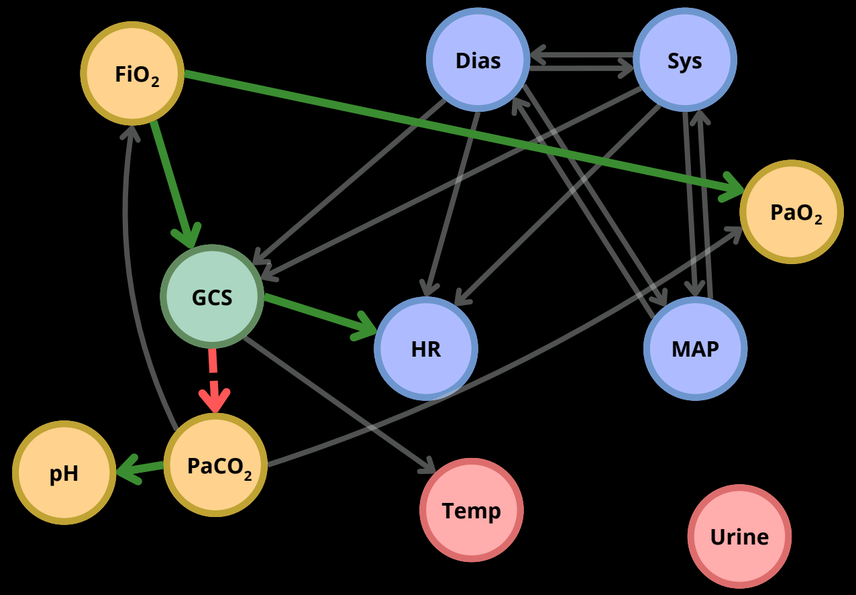


Cardiac Surgery (CSRU)

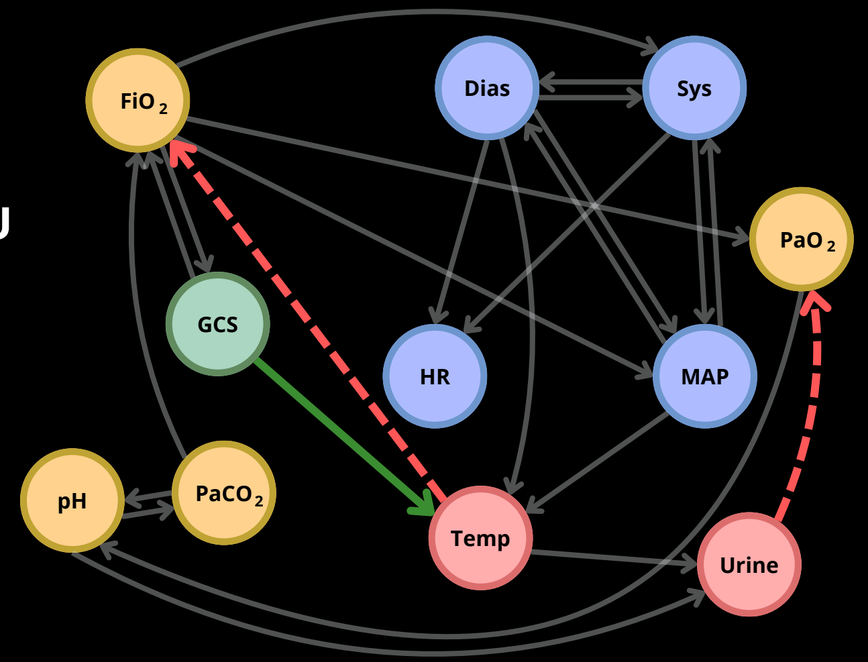


Results with different Standardization Techniques

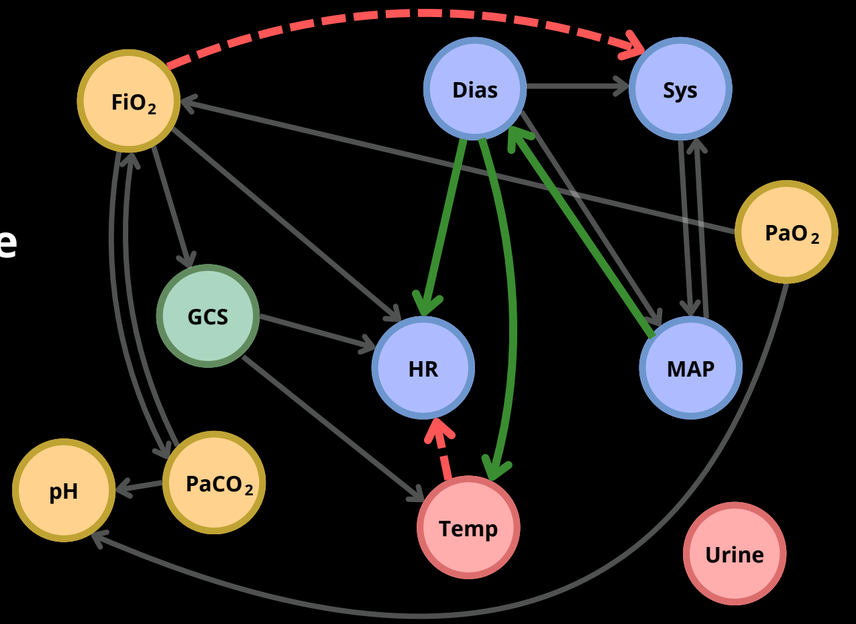
Medical ICU (MICU)



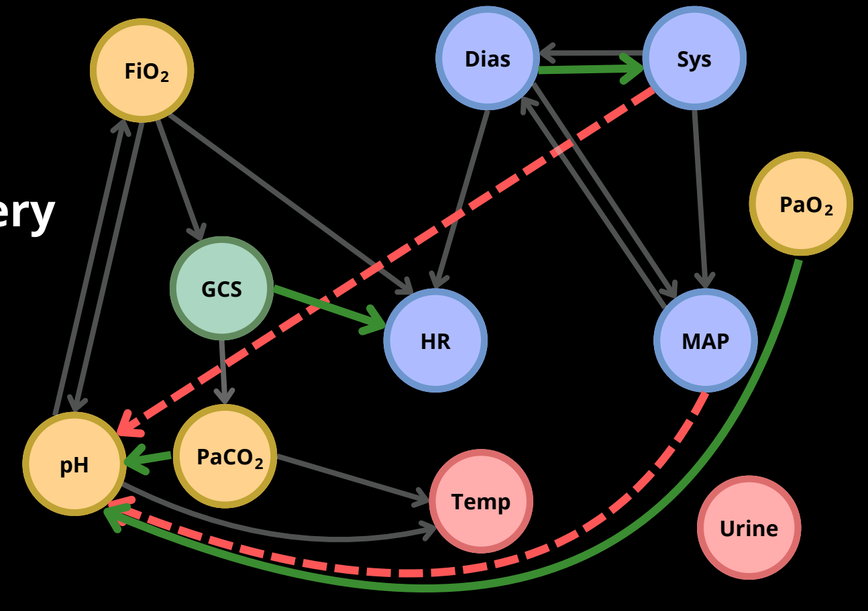
Surgical ICU (SICU)



Coronary Care (CCU)



Cardiac Surgery (CSRU)



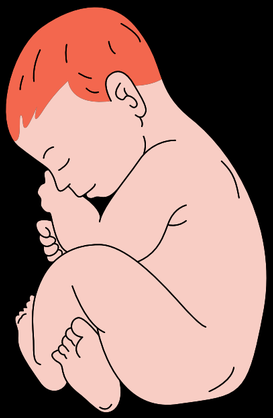
Global Standardization Only

Local Standardization Only

LUME-DBN Validation

SIMULATION STUDIES

ICU CHALLENGE DATA



PREMATURITY CLINICAL DATA

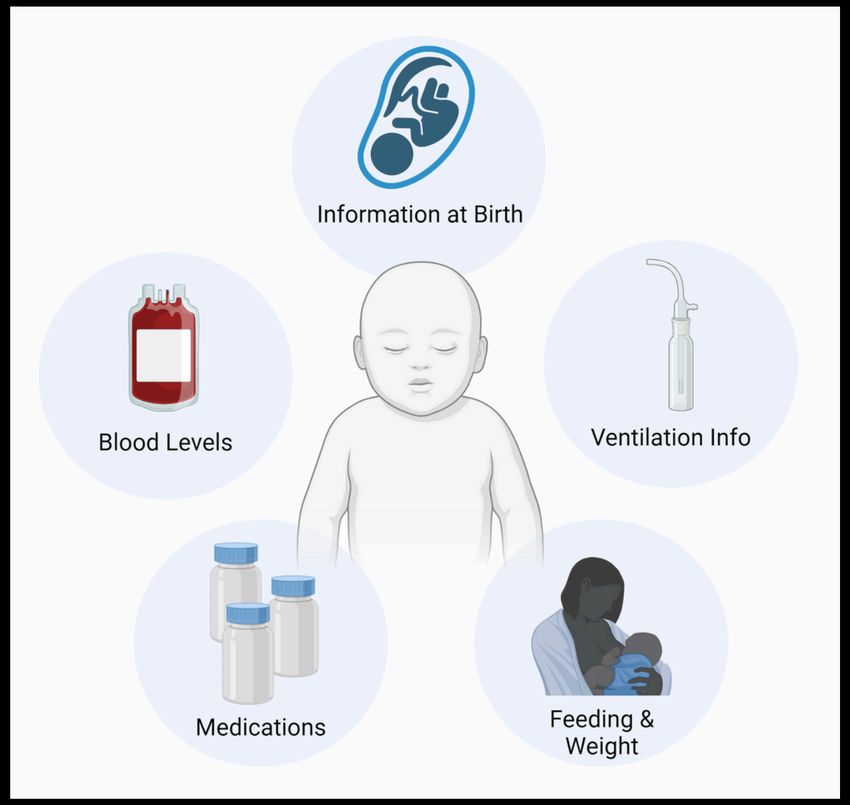
+ REAL-WORLD RETROSPECTIVE DATA

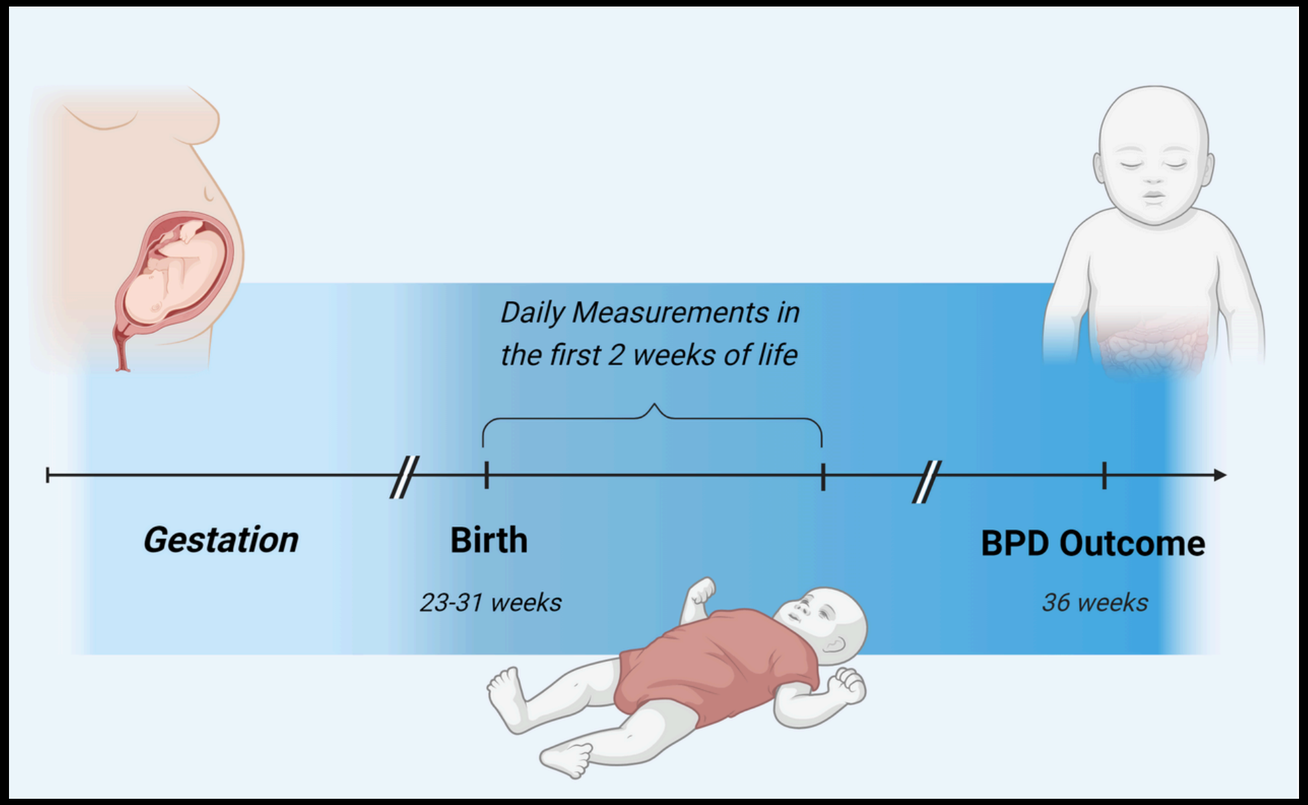
+ KNOWN DATA COLLECTION PROCEDURE

+ CLINICAL DOMAIN KNOWLEDGE

– NO REFERENCE STRUCTURE

A case study on Bronchopulmonary Dysplasia





A multisciplinary project in collaboration with the neonatologist of the TIN Mangiagalli with the aim of early detecting BPD and understanding respiratory dynamics in extremely and very preterm newborns

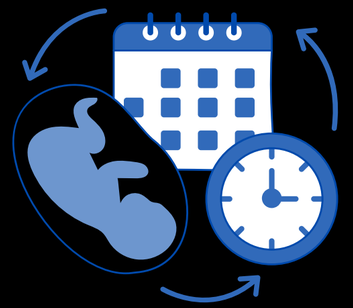
Data Preprocessing

DATA COLLECTION

- Patients born from 2017 to 2022 with gestational age (GA) < 32 weeks and birth weight <1500 g
- Data extraction from the Electronic Medical Records (EMR)

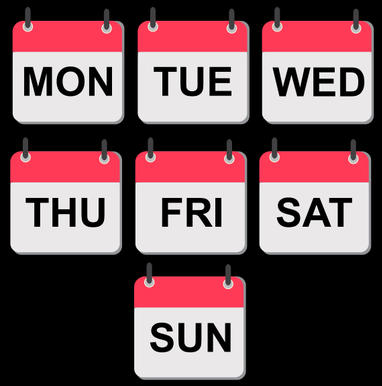
SAMPLE SELECTION

- Excluding patients with congenital malformations and outborns
- Ruling out patients not disposing of BPD outcomes at 36 weeks



DISCRETIZATION

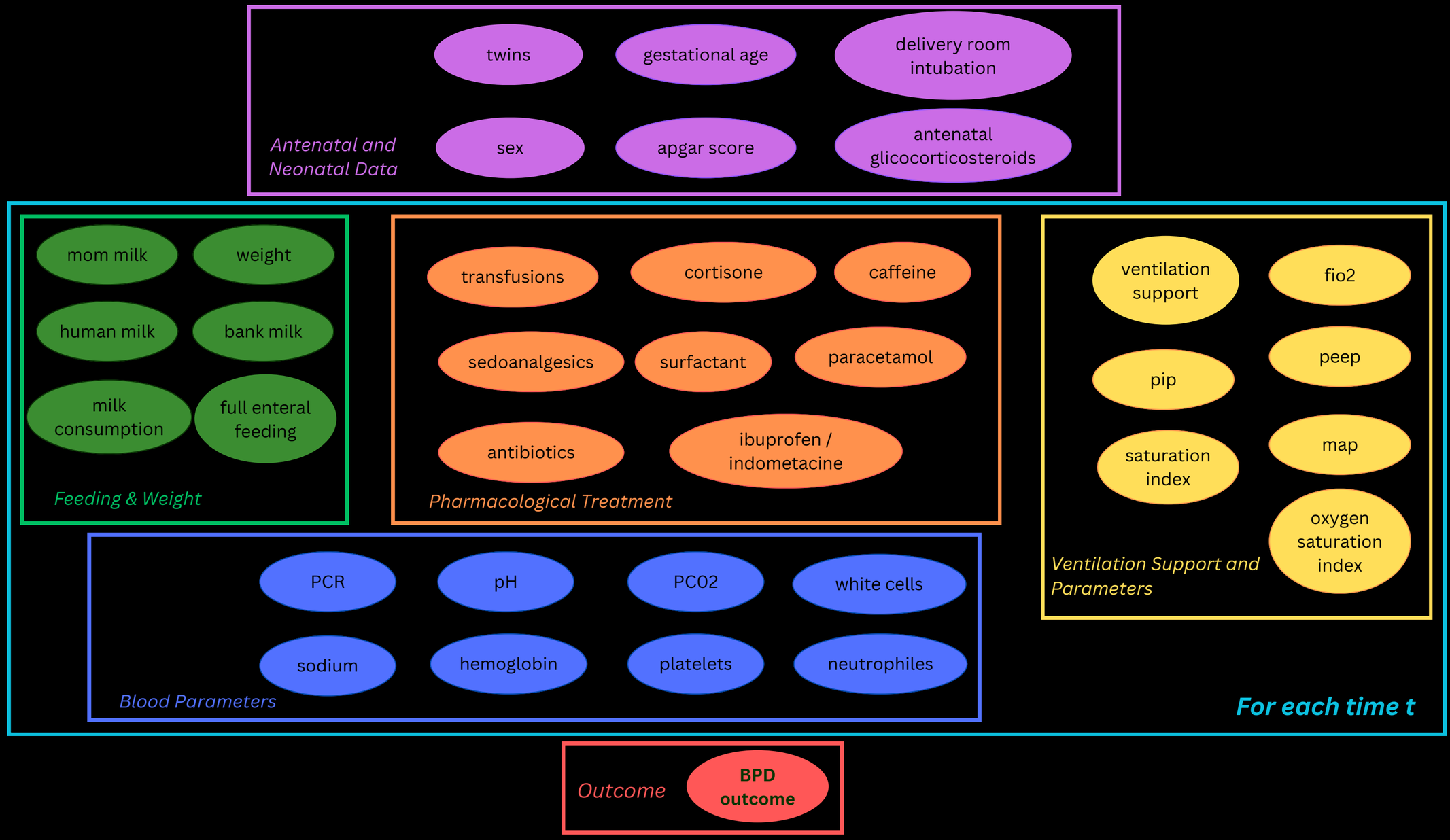
- Time discretization aggregating dynamic data on a daily basis (cumulative dosages, median ventilation and blood parameters)



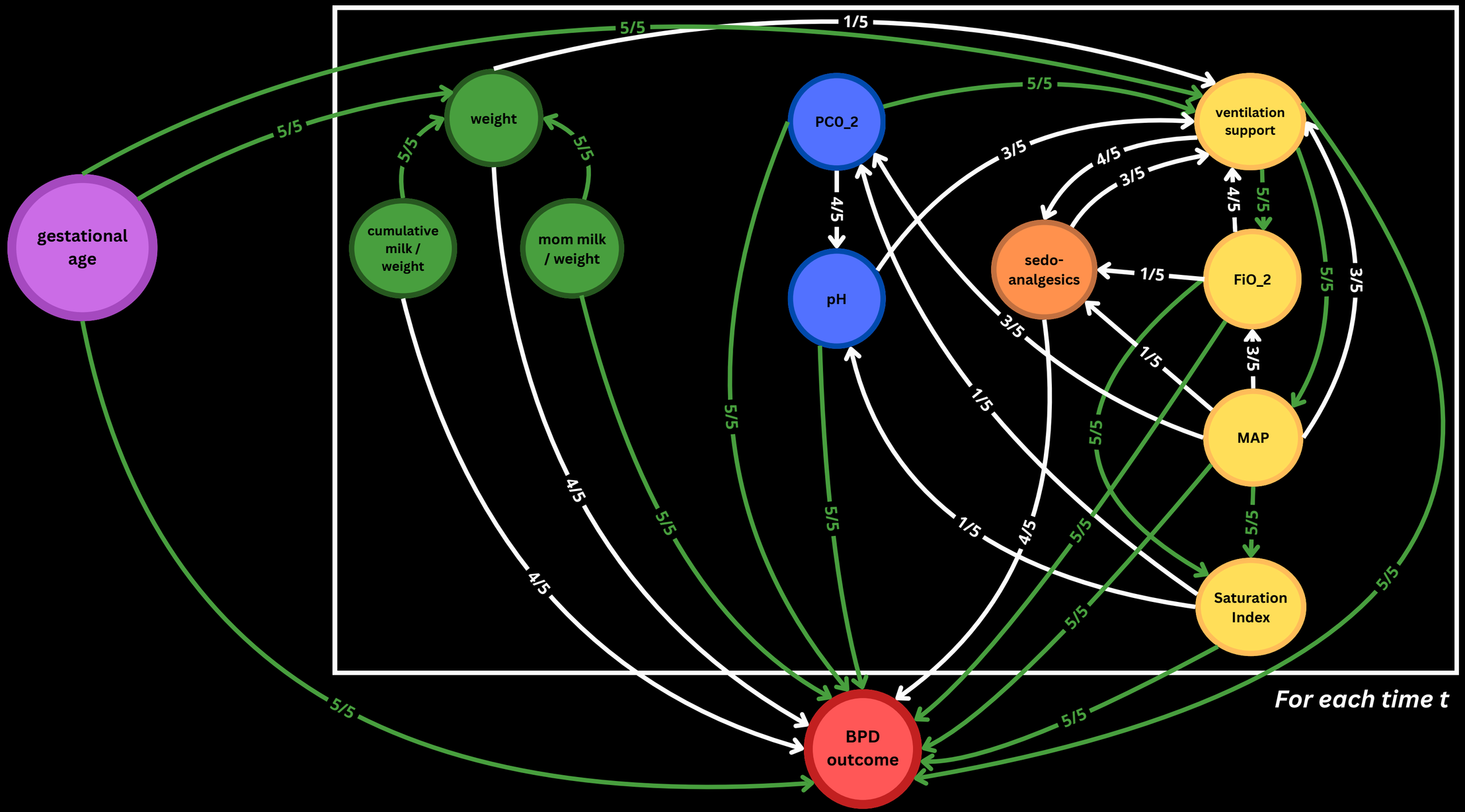
DATA CLEANING

- Validation of the features ranges based on laboratory thresholds
- The resulting dataset is comprised of mixed static and temporal data from 461 babies (287 healthy and 174 BPD)

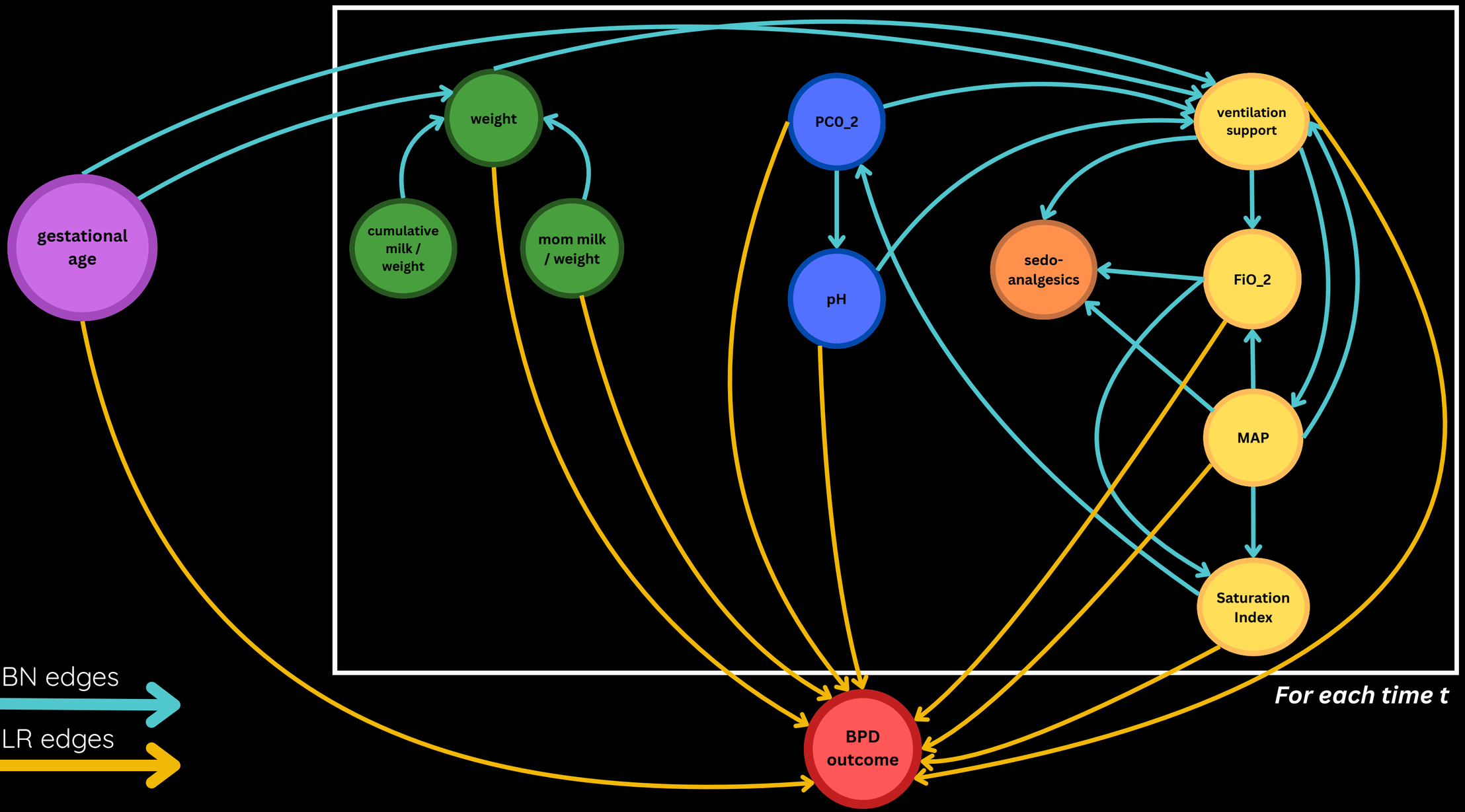
Incorporating prior structural knowledge



Prior informed DBN + Bayesian Logistic Regression (BLR) learning



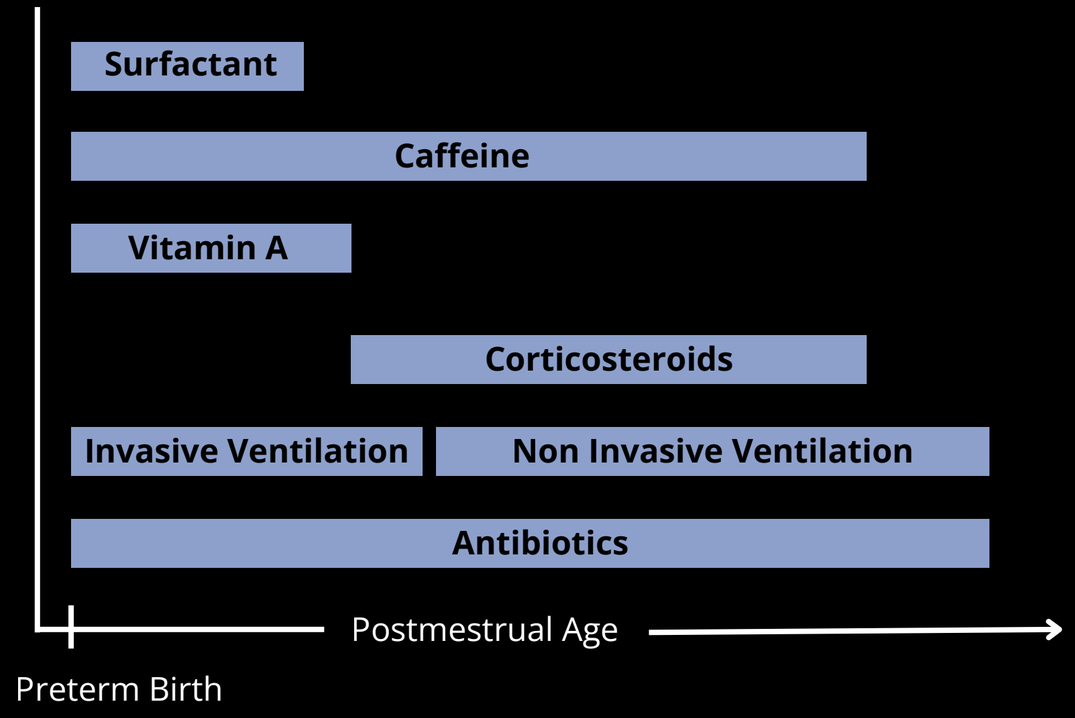
Expert-informed Learning of DBN + BLR



Why Standard (Homogeneous) DBNs could be suboptimal?

Temporal Shifts in System Dynamics

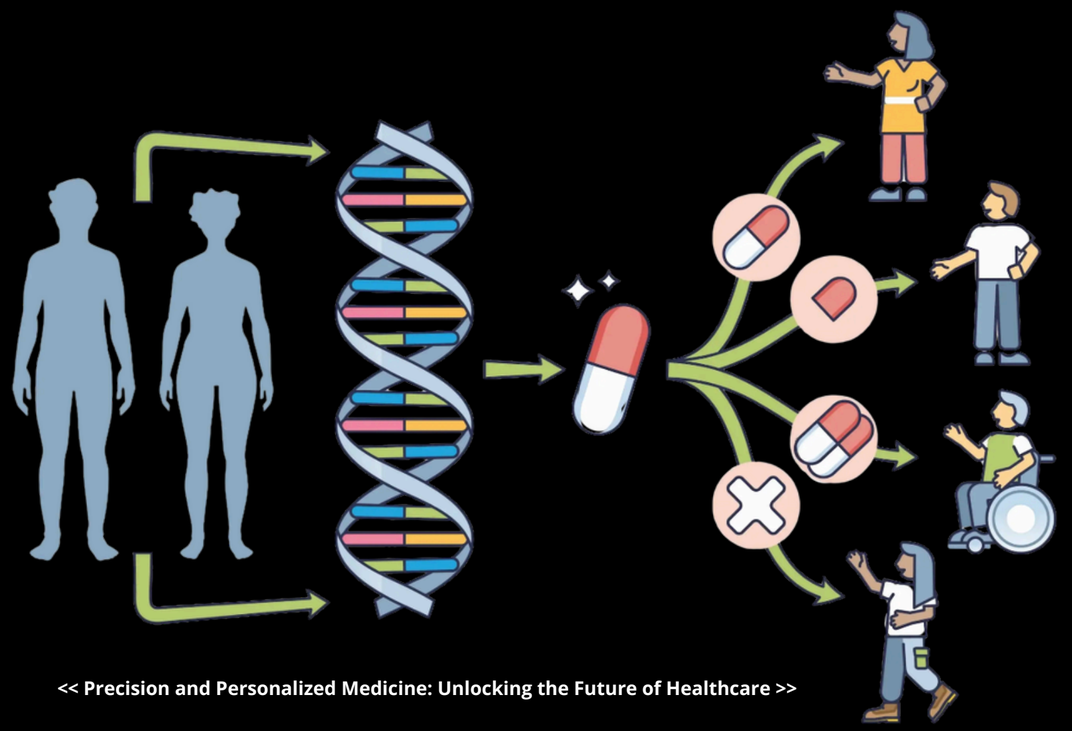
Unstable patients condition in the first few weeks of life



European standards for treatment adoption in premature newborns

Evolving temporal mechanism identification towards a better patient care over time

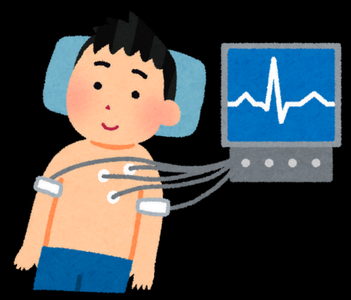
Deviation in treatment effects across patients groups



Treatment Effects changing from patient to patient

MEDICAL ICU (MICU)

SURGICAL ICU (SICU)



CORONARY CARE (CCU)

CARDIAC SURGERY (CSRU)

Distinct Groups of Patients with diverse conditions

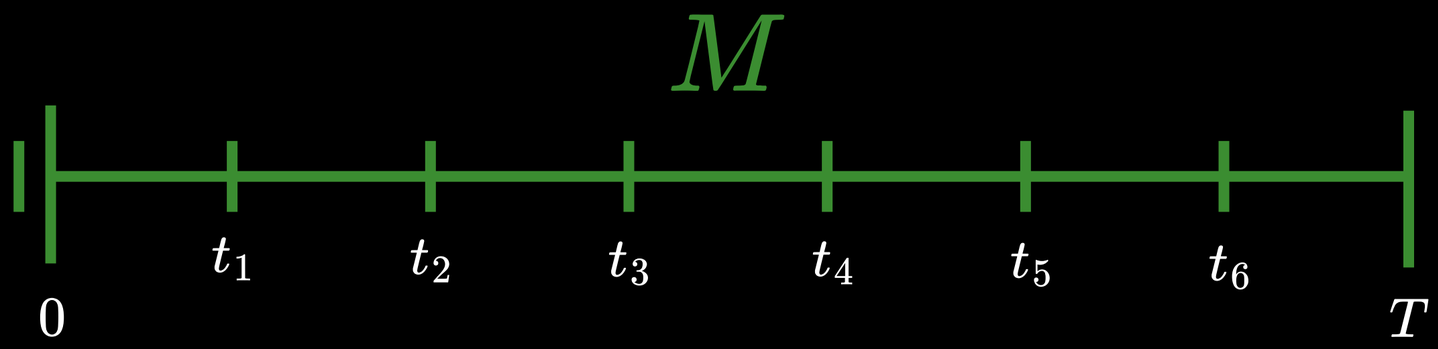
Varying dynamics identification across different groups towards personalize patient care

Non-stationary Temporal Models

Non-stationary Temporal Models

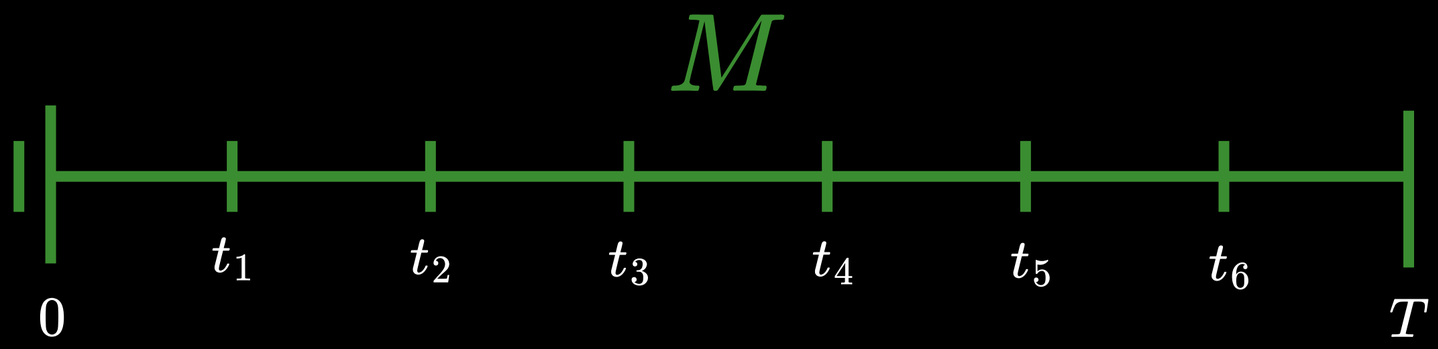
STATIONARY MODEL ACROSS

THE WHOLE PERIOD

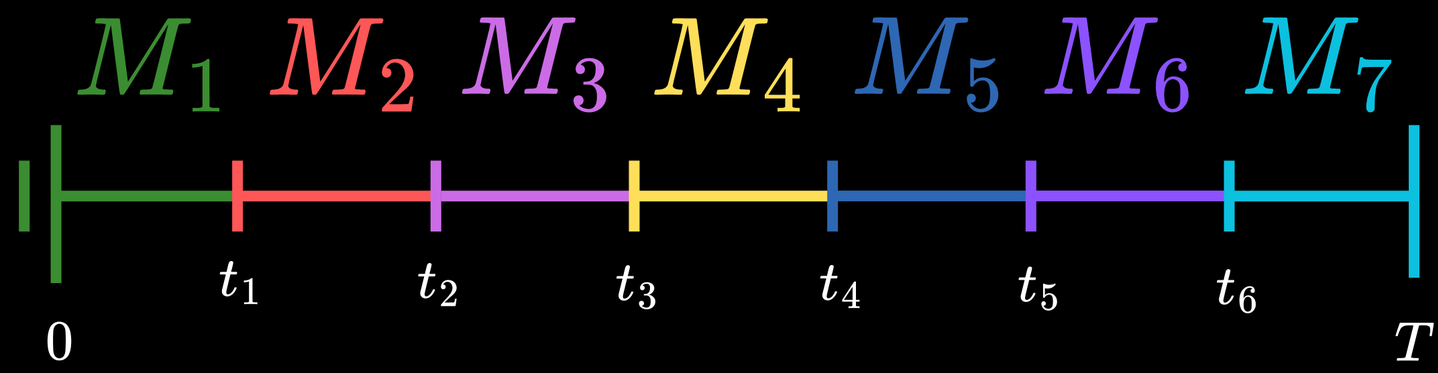


Non-stationary Temporal Models

STATIONARY MODEL ACROSS THE WHOLE PERIOD

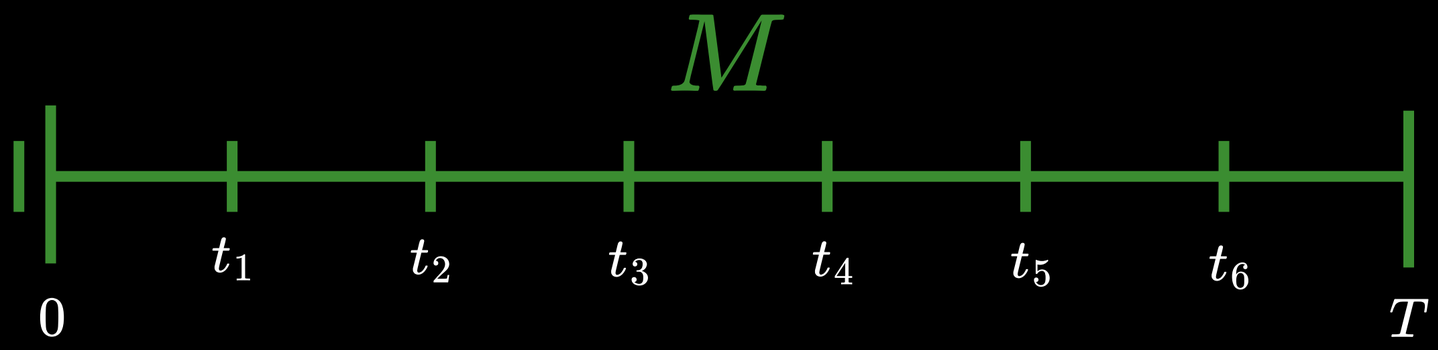


INDEDENDENT MODELS FOR EACH TIME FRAME

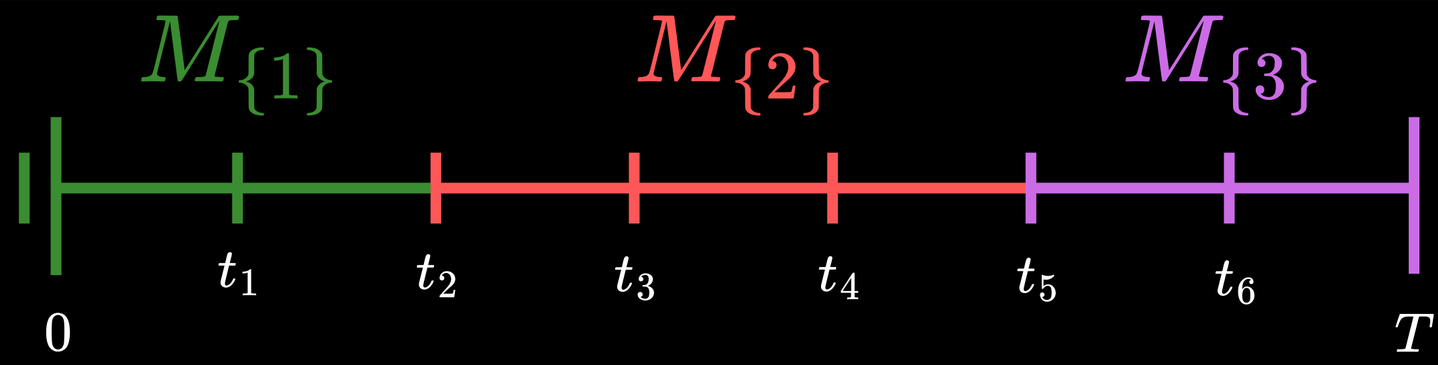


Non-stationary Temporal Models

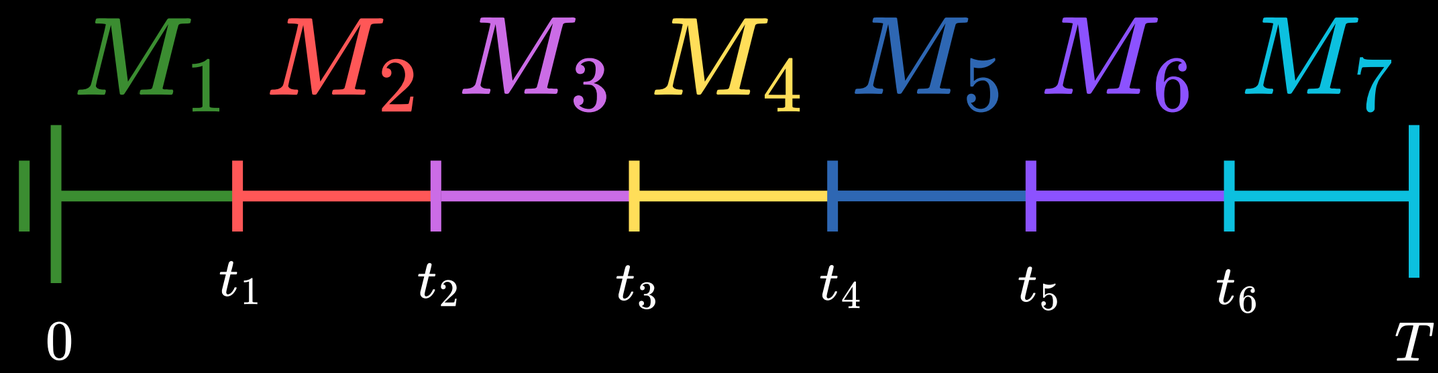
STATIONARY MODEL ACROSS
THE WHOLE PERIOD



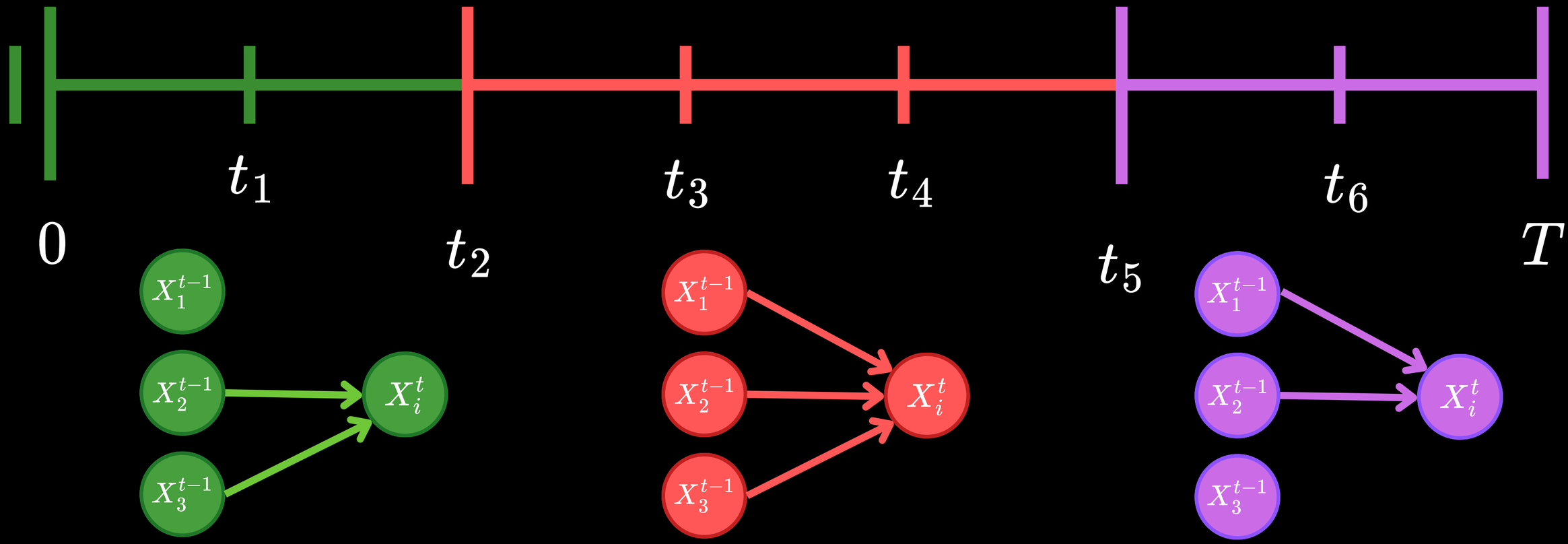
NON-STATIONARY MODEL
with a CHANGEPOINT PROCESS



INDEDENDENT MODELS FOR
EACH TIME FRAME



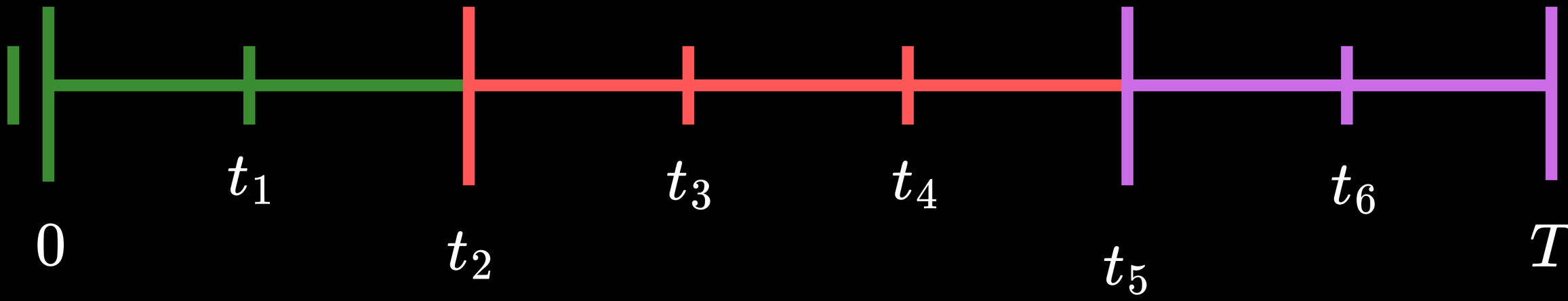
Non-stationary and Non-homogeneous DBNs



NON-STATIONARY DBNs

Structure Changes Over Time

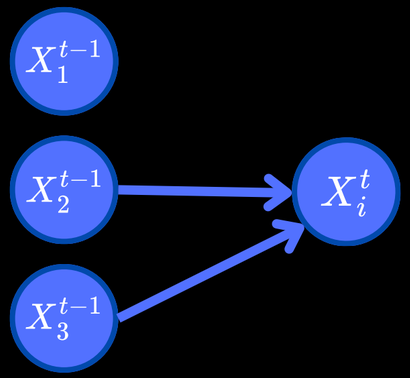
Non-stationary and Non-homogeneous DBNs



$$x_i^t = \beta_{0\{h_1\}} + \beta_{2\{h_1\}} x_2^{t-1} + \beta_{3\{h_1\}} x_3^{t-1}$$

$$x_i^t = \beta_{0\{h_3\}} + \beta_{2\{h_3\}} x_2^{t-1} + \beta_{3\{h_3\}} x_3^{t-1}$$

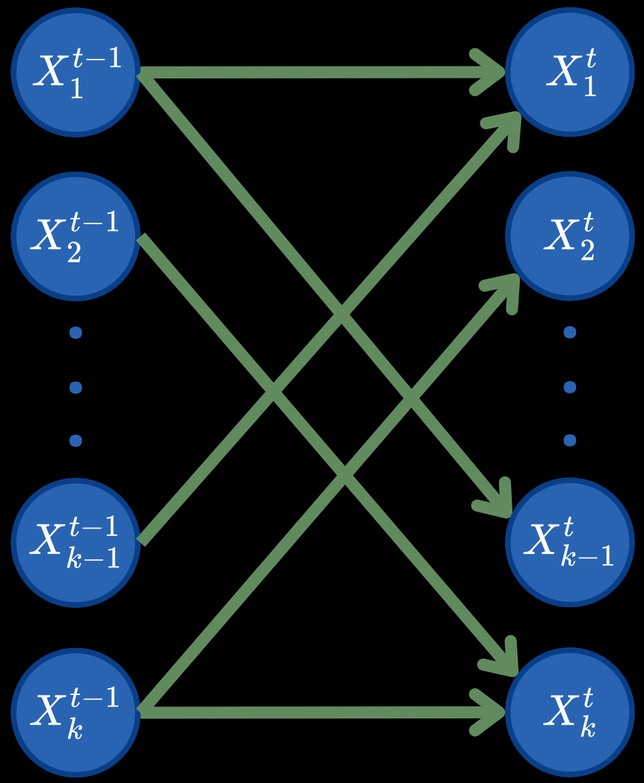
$$x_i^t = \beta_{0\{h_2\}} + \beta_{2\{h_2\}} x_2^{t-1} + \beta_{3\{h_2\}} x_3^{t-1}$$



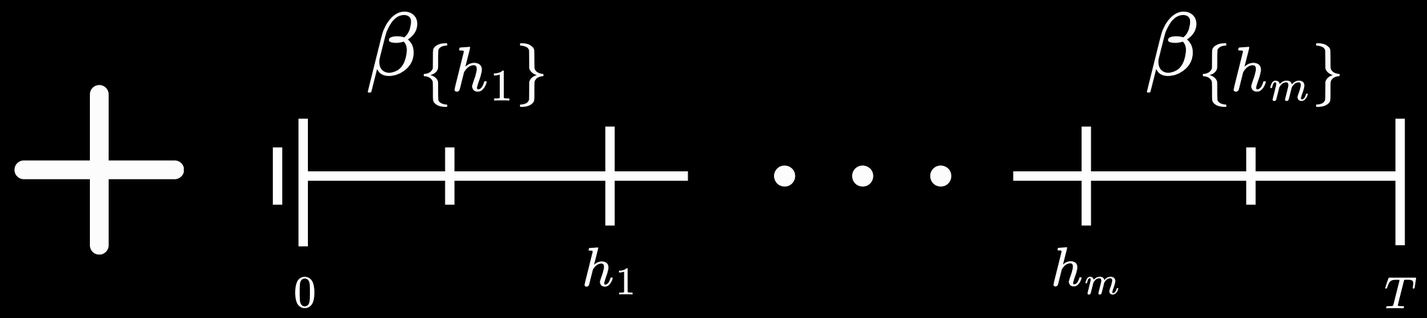
NON-HOMOGENEOUS DBNs

Structure is Fixed - Parameters Change Over Time

Non-homogeneous Dynamic Bayesian Networks



TRANSITION NETWORK

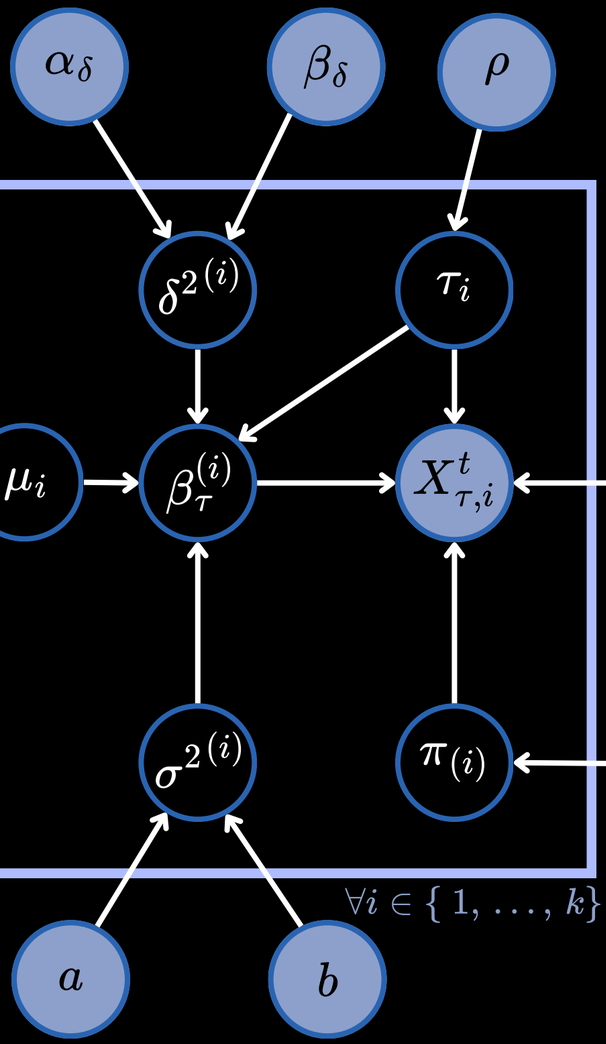


changepoints: $\{h_1, \ldots, h_m\}$

A SET OF CHANGEPOINTS

Learning NH-DBNs from Incomplete Data

Globally Coupled Non-Homogeneous DBNs



$\delta^{2(i)} \sim \text{Inv-Gamma}(\alpha_\delta, \beta_\delta)$

$\tau_i = \{t_1, \ldots, t_{|\tau_i|}\}$

$d_h := t_h - t_{h-1} \sim Geometric(\rho)$

$\mu_i \sim \mathcal{N}(\mu^{(\times)}, \sigma^{2(\times)})$

$X_{\tau,i}^t \mid (\mathcal{X}^{t-1}, \beta_\tau^{(i)}, \sigma^{2(i)}, \pi_{(i)}) \sim \mathcal{N}\left(\mathcal{X}_{[\pi_{(i)}],\tau}^{t-1} \beta_\tau^{(i)}, \sigma^{2(i)} I\right)$

${\beta_\tau}^{(i)} | (\sigma^{2(i)}, \delta^{2(i)}, \pi_{(i)}) \sim \mathcal{N}(\mu_i, \sigma^{2(i)} \delta^{2(i)} I)$

$|\pi_{(i)}| \sim \text{Poisson}(\lambda)$

$\sigma^{2(i)} \sim \text{Inv-Gamma}(a, b)$

LUME-NH Algorithm

LUME-NH is an extension of **LUME-DBN** for learning NH-**DBNs** from **Incomplete data.**

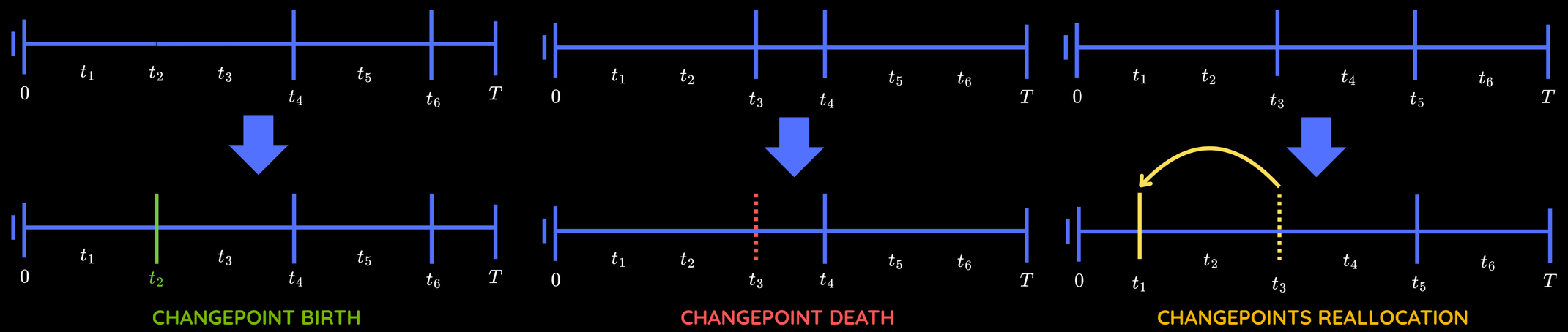
For each outcome variable

For each sample and time frame

Changepoint Set Update

Conditional on the current structure and missing values

Sampling a Move and Propose a new set of Changepoints



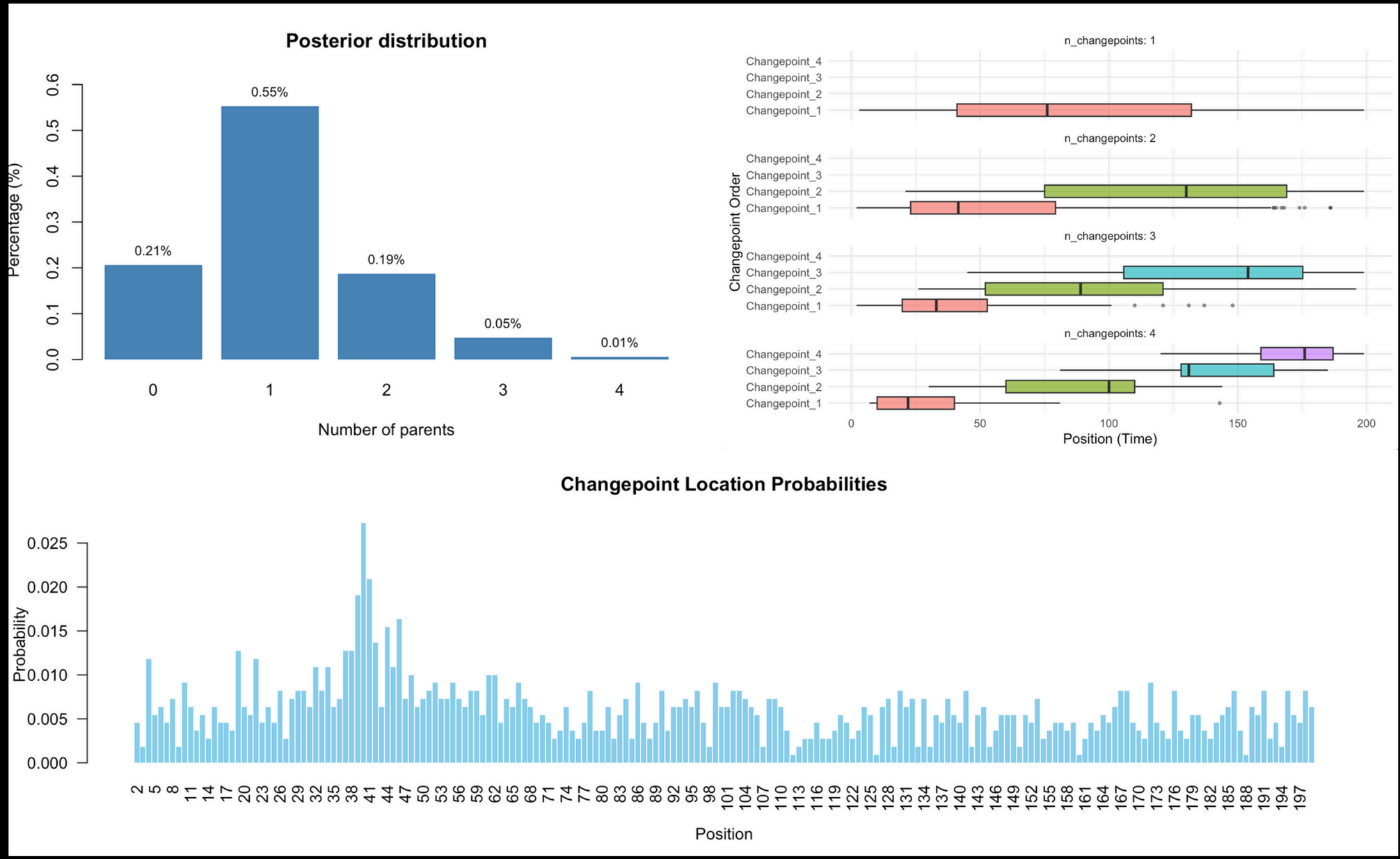
Sampling a New Global Mean Vector

$$\tau_\star \sim N(\mu^{\times\times}, \Sigma^{\times\times})$$

Changepoint Set Update Probability of Acceptance for $A > p \sim U(0,1)$

$$A([\tau, \mu] \to [\tau_\star, \mu_\star]) = \min\left\{1,\ \frac{p(Y \mid \tau_\star, \mu_\star, \pi, \delta^2)}{p(Y \mid \tau, \mu, \pi, \delta^2)} \cdot \frac{p(\tau_\star)}{p(\tau)} \cdot \frac{p(\mu_\star)}{p(\mu)} \cdot \frac{p(\mu|\sigma^2, \delta^2, \pi, \tau, Y)}{p(\mu_\star|\sigma^2, \delta^2, \pi, \tau_\star, Y)} \cdot HR\right\}$$

Posterior Distributions



Take Home Messages

NO MODEL FITS EVERYTHING

- In real-world settings, controlling all external factors is impracticable. Every model is therefore imperfect.
- Searching for better models matters, but acknowledging and quantifying uncertainty matters even more.

THE APPLICATION DRIVES EVERYTHING

- A strong theoretical setting is essential, but real problems emerge in practice.
- Analyze your data. Visualize it. Let what you see challenge, and try to break your assumptions.

EXPERT KNOWNLEDGE IS NOT OPTIONAL

- Purely data-driven approaches rarely succeed, especially with limited data.
- Expert insight comes from literature, experience, and scientific reasoning. Ignoring what is already known is not innovation. It’s information loss.

Thank You

UNIVERSITÀ DEGLI STUDI DI MILANO BICOCCA
MAD Lab
DISCO DIPARTIMENTO DI INFORMATICA SISTEMISTICA E COMUNICAZIONE
university of groningen
Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico
Anthem