Temporal Models

$X_i^{(t)}$ represents the instantiation of variable X_i at time t. X_i is a template variable

each "possible world" is a trajectory.
our goal is to represent a joint distribution over such trajectories

A dynamic system over the template variable X satisfies the Markov assumption if $\forall t \geq 0$

$$(X^{(t+1)} \perp X^{(0:t-1)} \mid X^{(t)})$$

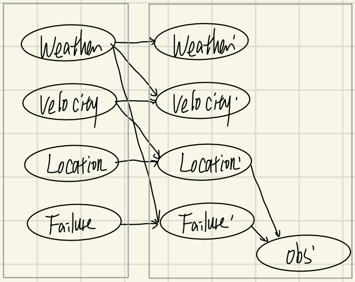
$$P(X^{(0:T)}) = P(X^{(0)}) \cdot \prod_{t=1}^{T} P(X^{(t)} \mid X^{(t-1)})$$

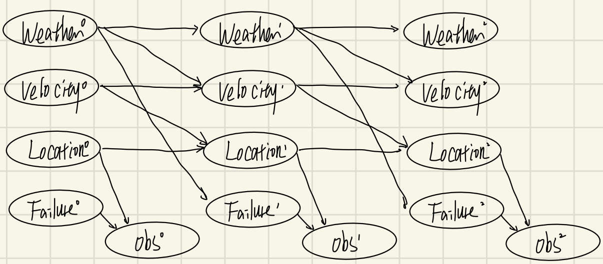
A Markovian dynamic system is stationary (invariant/homogeneous) if $P(X^{(t+1)} \mid X^{(t)})$ is same for all t. In this case, we can represent the process using a transition model $P(X' \mid X)$

$$P(X^{(t+1)} = \xi' \mid X^{(t)} = \xi) = P(X' = \xi' \mid X = \xi) \quad \forall t \geq 0$$

Dynamic Bayesian Network

A 2-time-slice Bayesian Network (2-TBN) for a process over X is a conditional Bayesian Network over X' given X_I, where $X_I \subseteq X$ is a set of interface variables





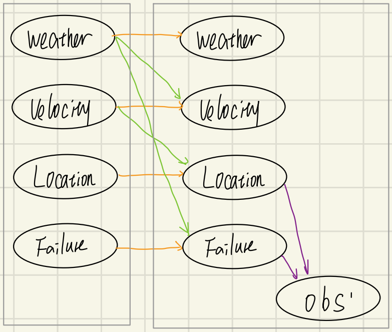
unrolled DBN

In a Conditional Bayesian Network, only X' have parents or CPDs
the interface variables X_I are those variables whose values at time t have a direct effect on variables at $t+1$
$\therefore$ only variables in X_I can be parents of variables in X'

A 2-time-slice Bayesian Network represents the conditional distribution

$$P(X' \mid X) = P(X' \mid X_I) = \prod_{i=1}^{n} P(X_i' \mid Pa(X_i'))$$

For each template variable X_i, the CPD $P(X_i' \mid Pa(X_i))$ is a template factor
it will be instantiated multiple times within the model, for multiple $X_i^{(t)}$ and their parents



$\rightrightarrows$: inter-time-slice edges

$\rightarrow$: edges of form $X \to X'$ are persistence edge
X is persistence variable

$\longrightarrow$: intra-time-slice
if the effect of one variable on the other is immediate
(much shorter than time granularity in the model)
there's a intra-time-slice

A dynamic Bayesian Network is a pair $\langle B_0, B_\to \rangle$, where B_0 is a Bayesian network over $X^{(0)}$ representing the initial distribution over states. And $B_\to$ is a 2-TBN for the process
For any desired time span $T \geq 0$, the distribution over $X^{(0:T)}$ is defined as a unrolled Bayesian network
(B_0 is the initial state, $B_\to$ is the transition)

· State - Observation Models

Partition variables into 2 Groups : X which is always hidden
O which is always observed

A State-observation model utilizes two independence assumptions:

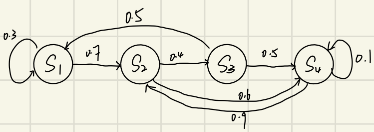
Markov assumption : $(X^{(t+1)} \perp X^{(0:t-1)} \mid X^{(t)})$

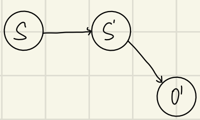
observation only depend on current state: $(O^{(t)} \perp X^{(0:t-1)}, X^{(t+1:\infty)} \mid X^{(t)})$

The probabilistic model consists 2 components:

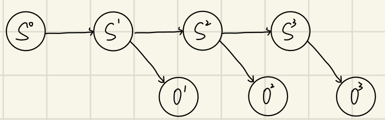
transition model $P(X' \mid X)$ and observation model $P(O \mid X)$

· State - Observation Models : Hidden Markov Models





2-TBN for a HMM



The unrolled DBN

A Hidden Markov Model (HMM) is a special of a simple DBN which has a sparse transition model

· State - Observation Models : Linear Dynamical System

A linear dynamical model represents a system of one or more real-valued variables that evolve linearly over time, with some Gaussian noise

A linear dynamical system (often Kalman filter) can be viewed as a dynamical Bayesian network where all variables are all continuous and all of the dependencies are linear Gaussian

$$P(X^{(t)} \mid X^{(t-1)}) = \mathcal{N}(A X^{(t-1)}; Q) \qquad A \in \mathbb{R}^{n \times n} \qquad Q \in \mathbb{R}^{n \times n}$$

$$P(O^{(t)} \mid X^{(t)}) = \mathcal{N}(H X^{(t)}; R) \qquad H \in \mathbb{R}^{m \times n} \qquad R \in \mathbb{R}^{m \times m}$$

A nonlinear variant of linear dynamical system (often called a extended Kalman filter) have nonlinear transition models and observation models

$$P(X^{(t)} \mid X^{(t-1)}) = f(X^{(t-1)}, U^{(t-1)})$$

$$P(O^{(t)} \mid X^{(t)}) = g(X^{(t)}, W^{(t)})$$

where f and g are deterministic nonlinear functions, and $U^{(t)}, W^{(t)}$ are Gaussian r.v.'s.

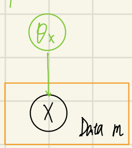
· Plate Models

In the plate formalism, object types are called types are called "plates"

The fact that multiple objects in the class share the same set of attributes and same probabilistic model is the basis for the use of the term "plate"

make explicit that all variables $X(d)$ are sampled from the CPD

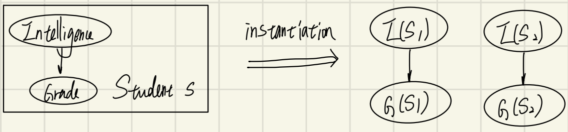
have a set of random variables $X(d)$ $(d \in D)$ that all have the same domain $Val(X)$ and are sampled from the same distribution



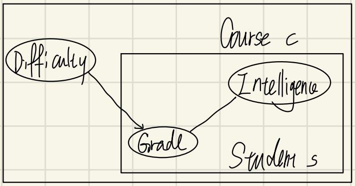
eg X is the template coin toss, $X(i)$ instantiate X as the ith toss

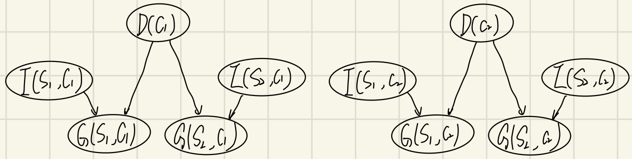
$X(1), X(2) \ldots \sim$ iid $Bernoulli(\theta_x)$

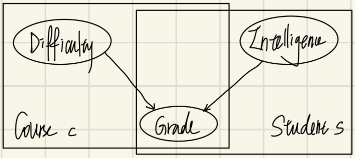
all $X(d)$ are generated from the same template

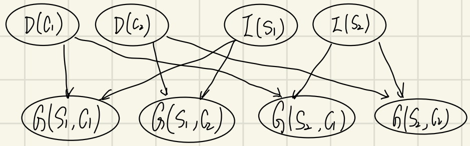


implicitly assume that $P(I(s))$ and $P(G(s) \mid I(s))$ is same for all s









∵ in blood type example, Genotype(child) depends on Genotype(father), plate models fail to capture this dependencies

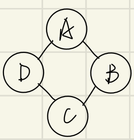
// Probabilistic Relational Model

- Undirected Templated Model

A relational Markov network $\mathcal{M}_{RMN}$ is defined in terms of a set of template features Λ where each template feature $\lambda \in \Lambda$ comprises

- A real-valued template feature f_λ whose arguments are $\mathcal{N}(\lambda) = \{A_1(U_1), \ldots, A_l(U_l)\}$
- A weight $w_\lambda \in \mathbb{R}$

$\alpha(\lambda) = \{U_i\}$



define $\begin{cases} \text{a binary predicate (relation)} & \text{Study-Pair}(S_1, S_2) \\ \text{a predicate (attribute)} & \text{Misconception}(S) \end{cases}$

define a template feature f_m over pairs $(\text{Misconception}(S_1), \text{Misconception}(S_2))$

$$f_m(\text{Misconception}(S_1), \text{Misconception}(S_2), \text{Study-Pair}(S_1, S_2)) = \begin{cases} 1 & \text{if } (\text{Study-Pair}(S_1, S_2) = \text{true}) \,\&\&\, (\text{Mis}(S_1) = \text{Mis}(S_2)) \\ 0 & \text{otherwise} \end{cases}$$

$\mathcal{N}(\lambda_m) = \{\text{Study-Pair}(S_1, S_2), \text{MisCon}(S_1), \text{MisCon}(S_2)\}$

$\alpha(\lambda_m) = \{S_1, S_2\}$

Given a Markov Network $\mathcal{M}$ and an object skeleton $\mathcal{K}$, define a ground Gibbs distribution $P_{\mathcal{K}}^{\mathcal{M}}$

The variables in the Network

$P_{\mathcal{K}}^{\mathcal{M}}$ contains a term $\exp(w_\lambda \cdot f_\lambda(\gamma))$ for each template $\lambda \in \Lambda$ and each assignment