Temporal Models

www.cs.wisc.edu/~page/cs760/

Goals for the lecture

(with thanks to Irene Ong, Stuart Russell and Peter Norvig, Jakob Rasmussen, Jeremy Weiss, Yujia Bao, Charles Kuang, Peggy Peissig, and Becca Willett)

you should understand the following concepts

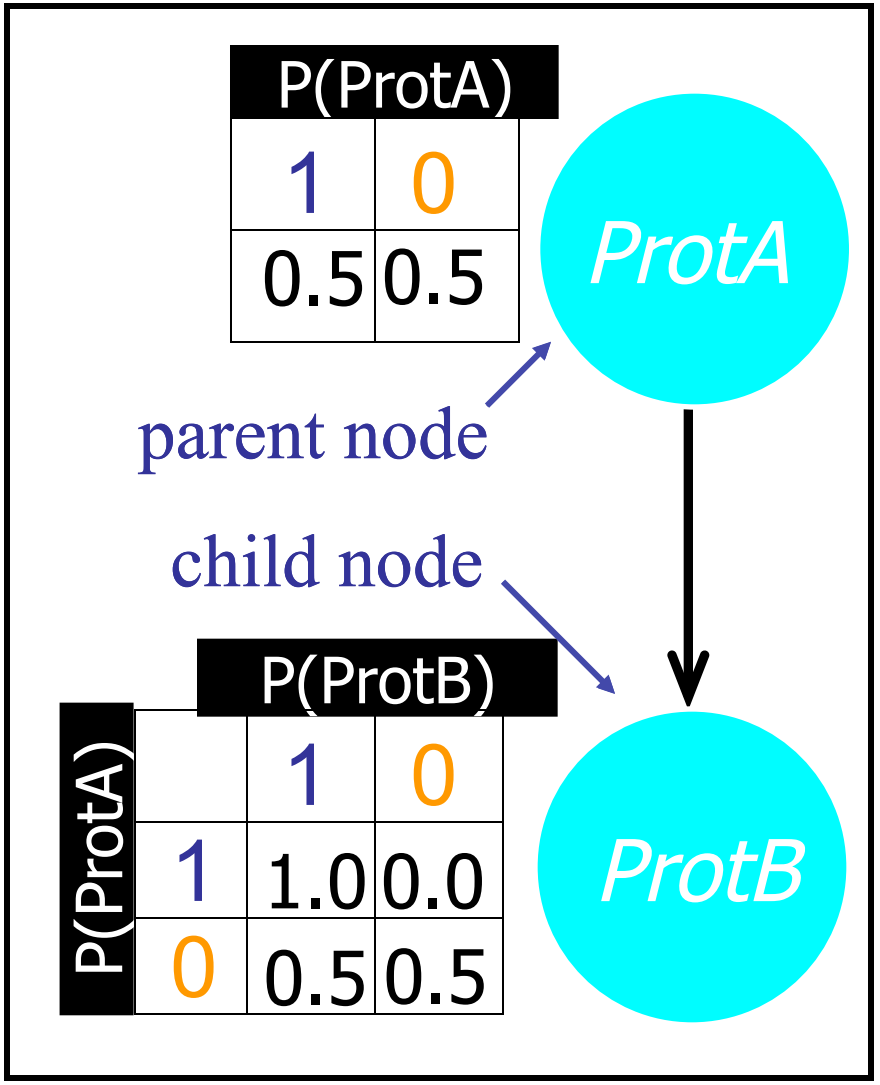
- dynamic Bayes nets (DBNs)
- continuous-time Bayes nets (CTBNs)
- point process models
- piecewise constant conditional intensity models (PCIMs)
- multiplicative forest point processes (MFPPs)
- Hawkes processes

Bayesian Network (BN)

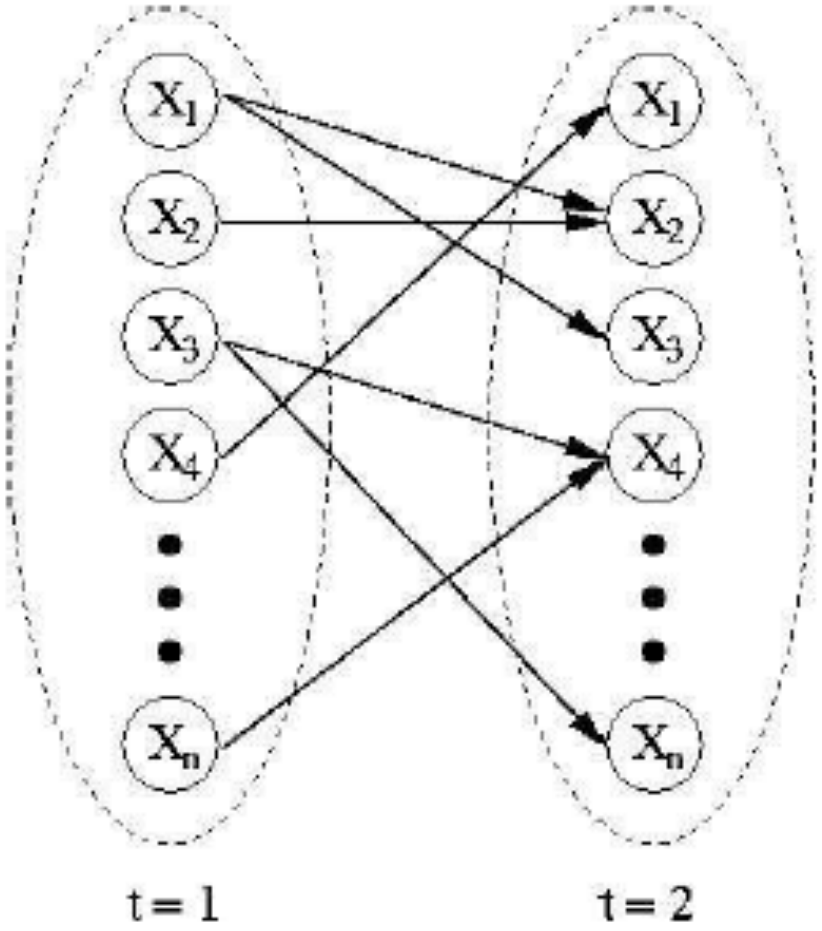
Data:

	ProtA	*ProtB*
Expt1	1	1
Expt2	1	1
Expt3	0	0
Expt4	0	1

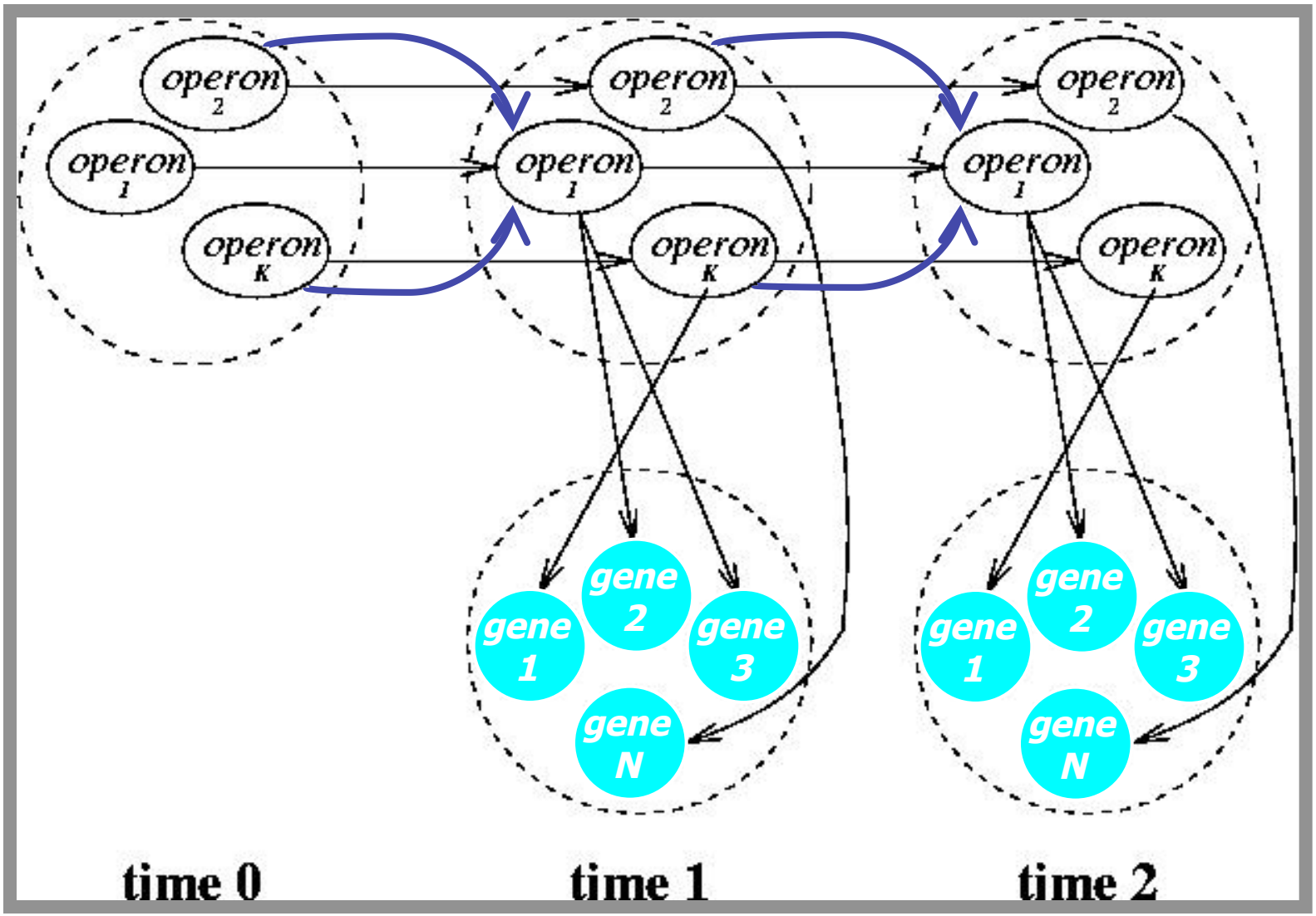
Friedman *et al.* (2000)



Dynamic Bayesian Network (DBN)



Example DBN [Ong et al., ISMB'02]



Another Example [Russell & Norvig]

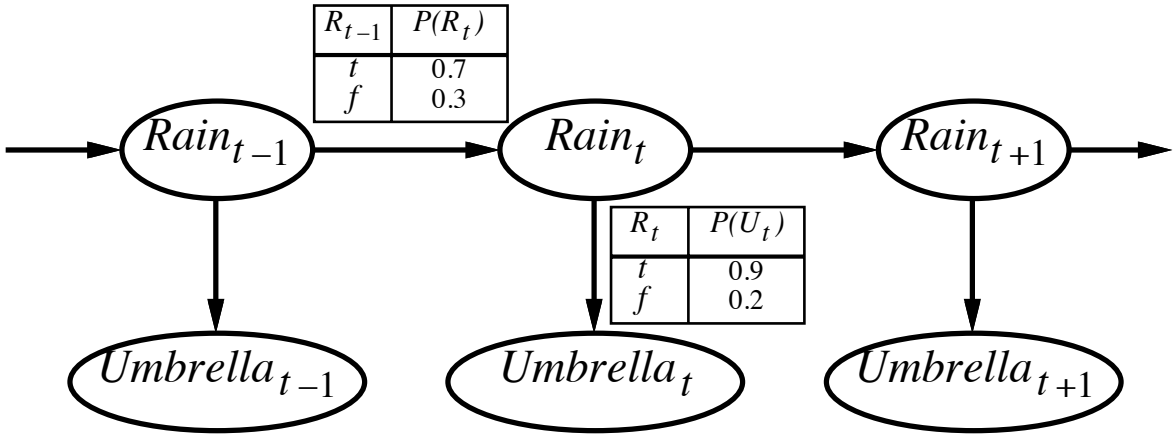


Figure 17.2 Bayesian network structure and conditional distributions describing the umbrella world.

Unrolling DBN [Russell & Norvig]

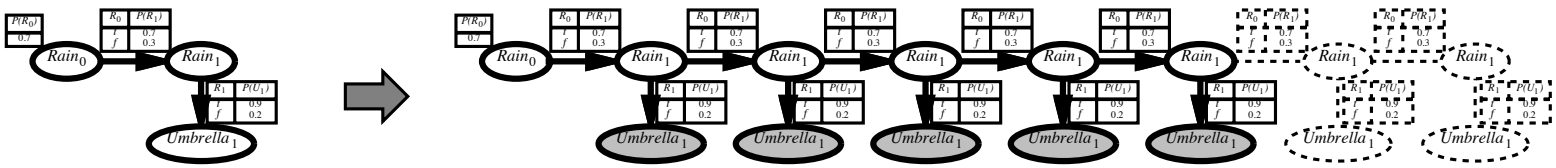
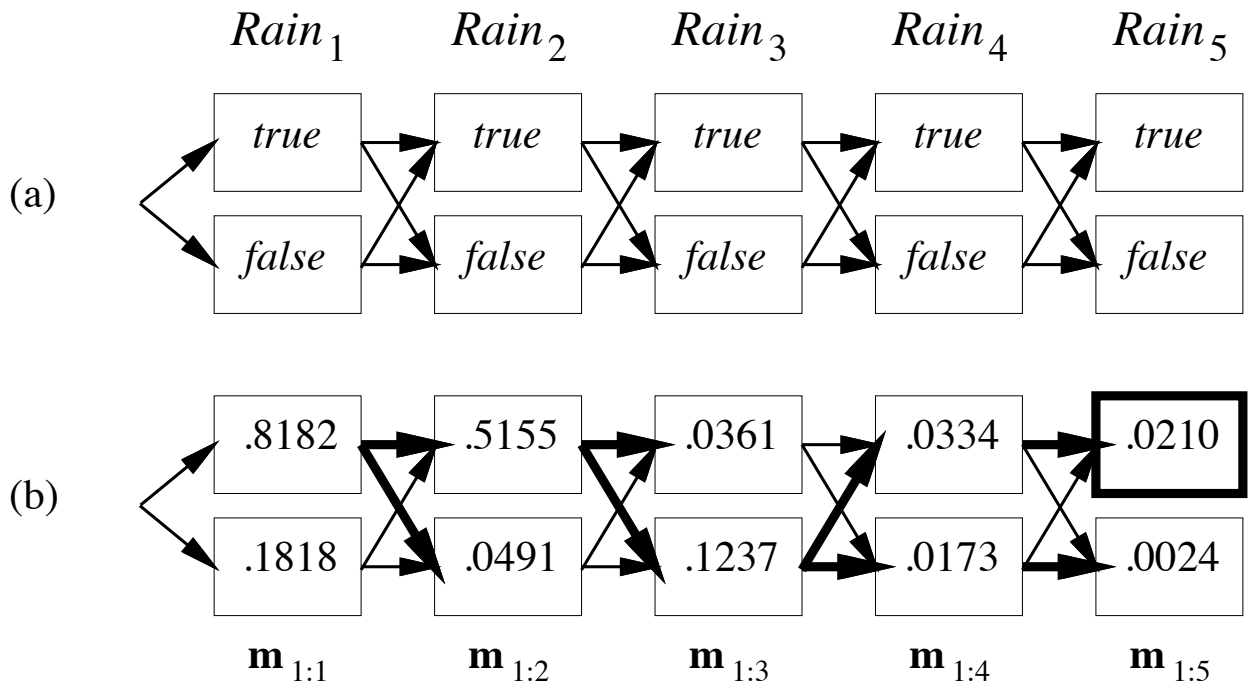


Figure 17.14 Unrolling a dynamic Bayesian network: slices are replicated to accommodate the observation sequence (shaded nodes). Further slices have no effect on inferences within the observation period.

DBN Algorithms

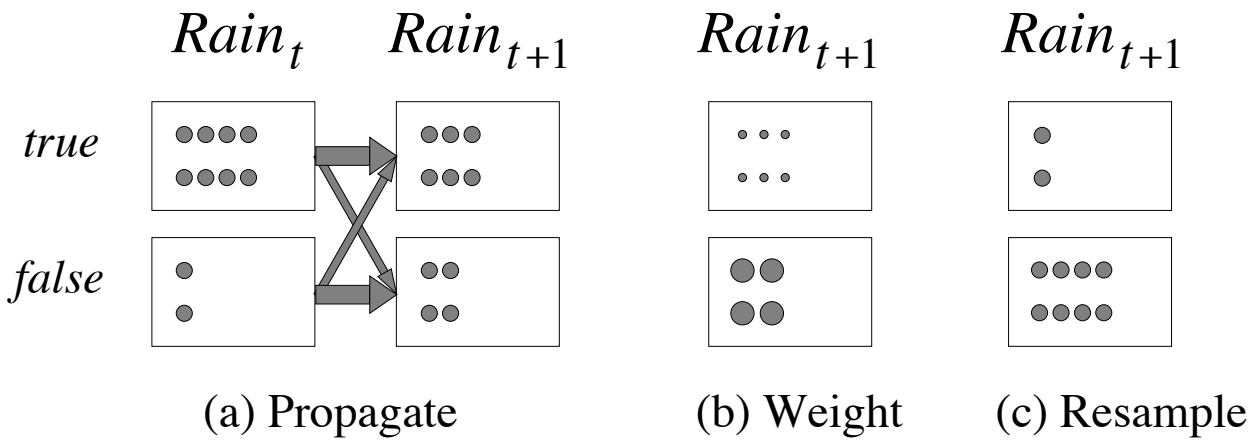
- Forward Algorithm: dynamic programming (equivalent to variable elimination in this case) to compute probability of a future variable
- Backward Algorithm: past variable
- Forward-Backward: compute probability of a state
- Viterbi: compute most probable trajectory (use max rather than sum over hidden variables)
- Baum-Welch: EM algorithm (over hidden state variables) to learn parameters
- Structure learning generally by greedy hill-climbing

Viterbi [Russell & Norvig]



Particle Filtering

- Choose number of particles N, and sample N particles
- At each subsequent time, from each particle, use transition to sample value at next time step
- Weight particles by likelihood given the evidence
- Resample N particles from this new, weighted distribution



Special Cases of DBNs

- Hidden Markov Models (HMMs)
  - One state variable
  - One emit variable
  - DBNs are factored HMMs

- Kalman Filters: special (original) case of DBNs where
  - CPDs are multivariate Gaussians
  - Linear dependence on parents, Gaussian noise
  - Transition (left-to-right) and Sensing (top-to-bottom) distributions
  - Developed to track objects by radar given observations every 10 seconds (or some such time interval)

Irregular Temporal Data Example: Clinical Data

60 yo M presents with sudden onset of substernal, heavy chest pain 45 minutes ago that radiates to left arm and neck. +dyspnea, +diaphoresis, +nausea.

History:
hypertension,
hyperlipidemia,
smoking

60 yo M presents with sudden onset of substernal, heavy chest pain 45 minutes ago that radiates to left arm and neck. +dyspnea, +diaphoresis, +nausea.

History:
hypertension,
hyperlipidemia,
smoking

Workup: ECG, cardiac enzymes, chest XR, CBC + electrolytes, echo

Myocardial Infarction (MI, heart attack)

60 yo M presents with sudden onset of substernal, heavy chest pain 45 minutes ago that radiates to left arm and neck. +dyspnea, +diaphoresis, +nausea.

History:
hypertension,
hyperlipidemia,
smoking

Workup: ECG, cardiac enzymes, chest XR, CBC + electrolytes, echo

Myocardial Infarction (MI, heart attack)

60 yo M presents with sudden onset of substernal, heavy chest pain 45 minutes ago that radiates to left arm and neck. +dyspnea, +diaphoresis, +nausea.

History:
hypertension,
hyperlipidemia,
smoking

Framingham risk score for men:
10pt = 60 yo
1 pt = hypertension
2 pt = hyperlipidemia
1 pt = smoking

14 pt → 16%
10-year risk of cardiovascular disease
(Wilson et al, *Circulation* 1998)

60 yo M presents with sudden onset of substernal, heavy chest pain 45 minutes ago that radiates to left arm and neck. +dyspnea, +diaphoresis, +nausea.

History:
hypertension,
hyperlipidemia,
smoking

Framingham risk score for men:
10pt = 60 yo
1 pt = hypertension
2 pt = hyperlipidemia
1 pt = smoking

14 pt → 16%
10-year risk of cardiovascular disease
(Wilson et al, *Circulation* 1998)

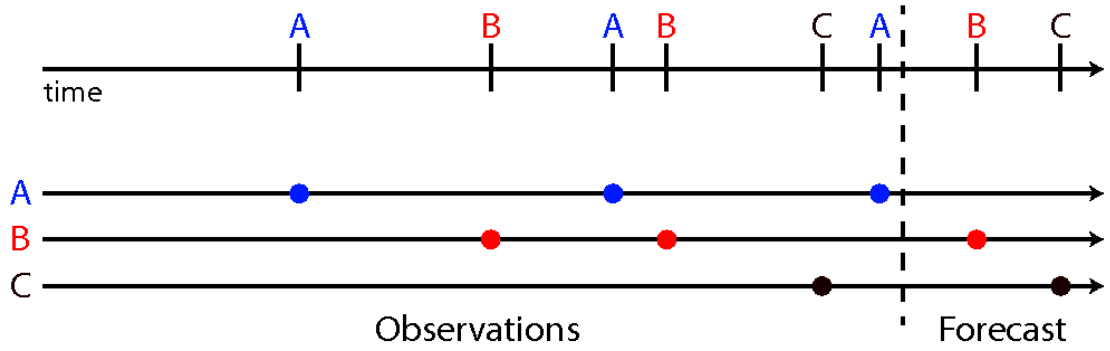
5-year risk of MI?
k-year risk of X?

60 yo M presents with sudden onset of substernal, heavy chest pain 45 minutes ago that radiates to left arm and neck. +dyspnea, +diaphoresis, +nausea.

History:
hypertension,
hyperlipidemia,
smoking

Framingham risk score for men:
10pt = 60 yo
1 pt = hypertension
2 pt = hyperlipidemia
1 pt = smoking

14 pt → 16%
10-year risk of cardiovascular disease
(Wilson et al, *Circulation* 1998)



5-year risk of MI?
k-year risk of X?

Timeline Representations

Continuous-time, discrete-state, with piecewise-constant transition rates

Point process: piecewise-continuous conditional intensity model (PCIM)
(Gunawardana et al., NIPS 2011)

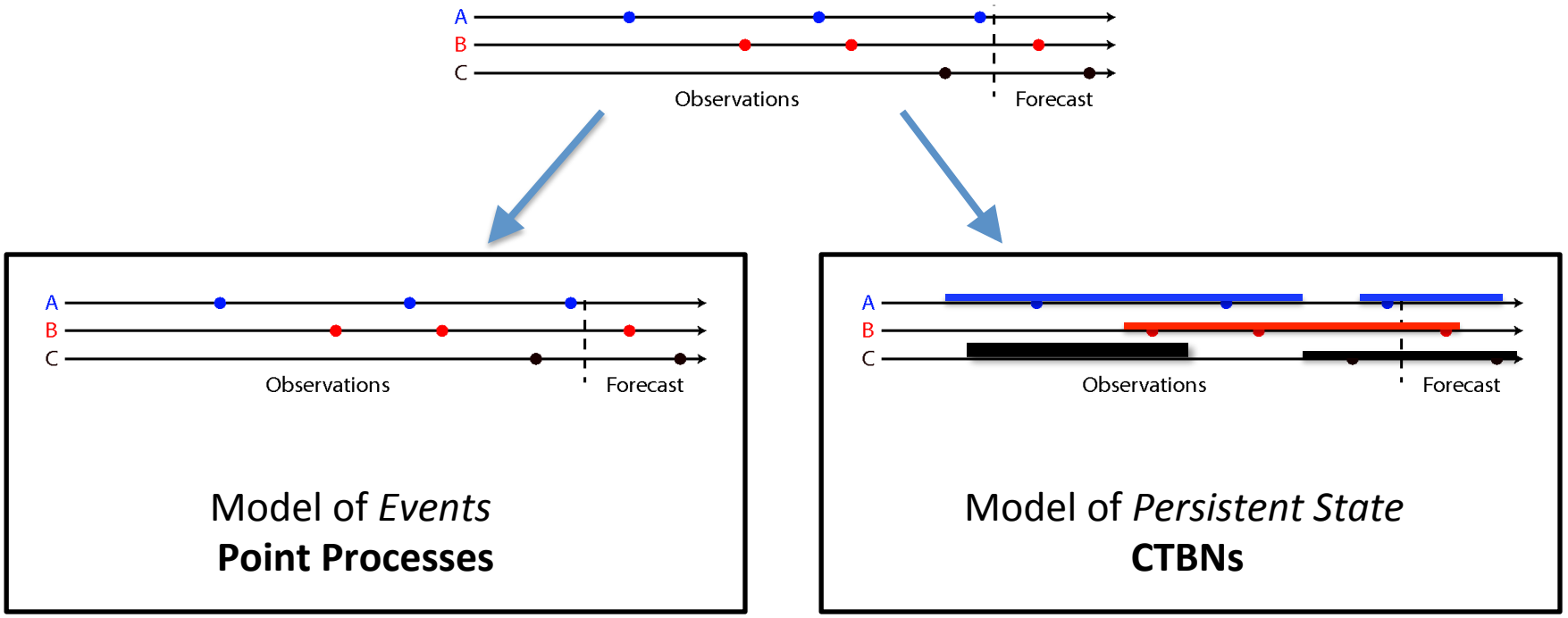
Continuous-time Bayesian networks (CTBNs) (Nodelman et al, UAI 2002)

Timeline Representations

Continuous-time, discrete-state, with piecewise-constant transition rates

Point process: piecewise-continuous conditional intensity model (PCIM)
(Gunawardana et al., NIPS 2011)

Continuous-time Bayesian networks (CTBNs) (Nodelman et al, UAI 2002)



Intensity Modeling

Event types ***l*** in ***L***
Trajectory ***x***: a sequence of time event pairs $(t,l)_i$
Rate function $\lambda(t|h)$ for {PCIM: events, CTBN: transitions}

Intensity Modeling

Event types l in L

Trajectory x: a sequence of time event pairs $(t,l)_i$

Rate function $\lambda(t|h)$ for {PCIM: events, CTBN: transitions}

Assumption: λ piece-wise constant

Dependency: {PCIM: basis states s in S, CTBN: variable states X}

states s in Σ_l, Σ_l mapping from x to S

e.g. PCIM: λ_a depends on event b in [t-1,t)

e.g. CTBN: λ_a depends on $B=b$

Intensity Modeling

Event types ***l*** in ***L***
Trajectory ***x***: a sequence of time event pairs $(t,l)_i$
Rate function $\lambda(t|h)$ for {PCIM: events, CTBN: transitions}
Assumption: λ piece-wise constant
Dependency: {PCIM: basis states ***s*** in ***S***, CTBN: variable states ***X***}
states ***s*** in Σ_l, Σ_l mapping from ***x*** to ***S***
e.g. PCIM: λ_a depends on event ***b*** in [t-1,t)
e.g. CTBN: λ_a depends on ***B=b***
Likelihood:

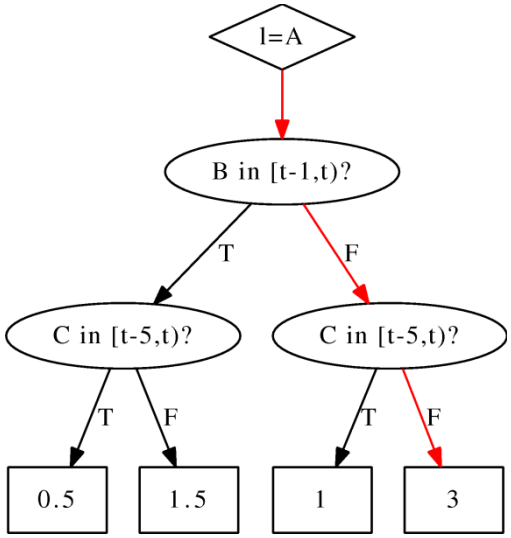
$$p(x|S,\theta) = \prod_{l\in\mathcal{L}} \prod_{s\in\Sigma_l} \lambda_{ls}^{M_{ls}(x)} e^{-\lambda_{ls} T_{ls}(x)}$$

M_{ls} : count of ***l*** given ***s***
T_{ls} : cumulative duration until ***l*** given ***s***

Point Process

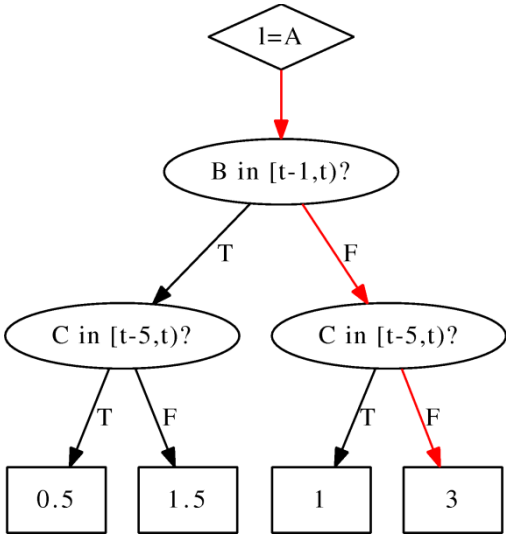
a.k.a., Piecewise-continuous Conditional Intensity Model (PCIM)

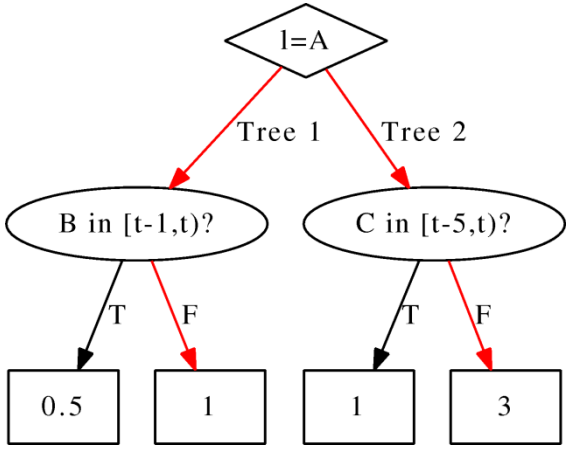
Represent dependencies with trees (Gunawardana et al, NIPS 2011)



Multiplicative forests

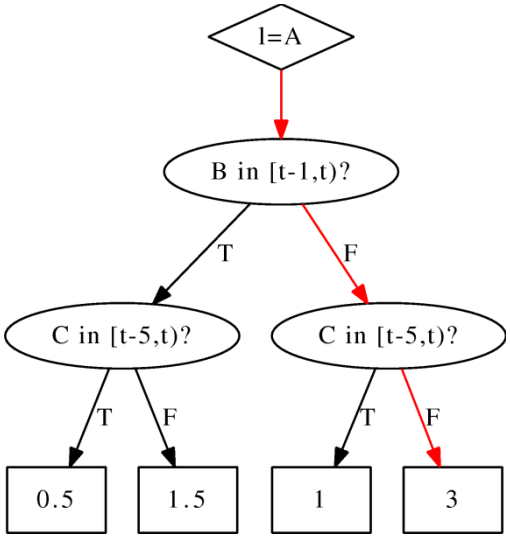
Represent dependencies with ~~trees~~ **forests**

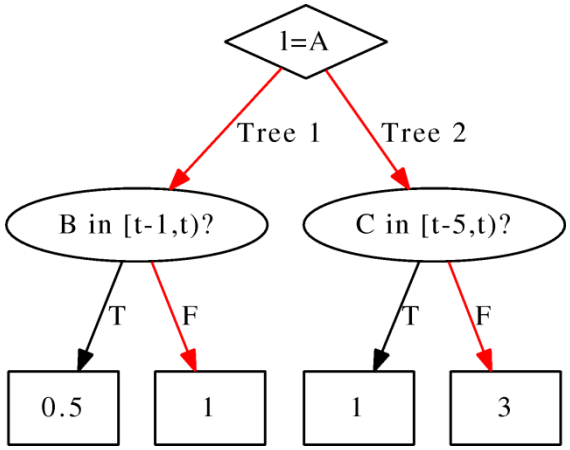




Multiplicative forests

Represent dependencies with ~~trees~~ **forests**



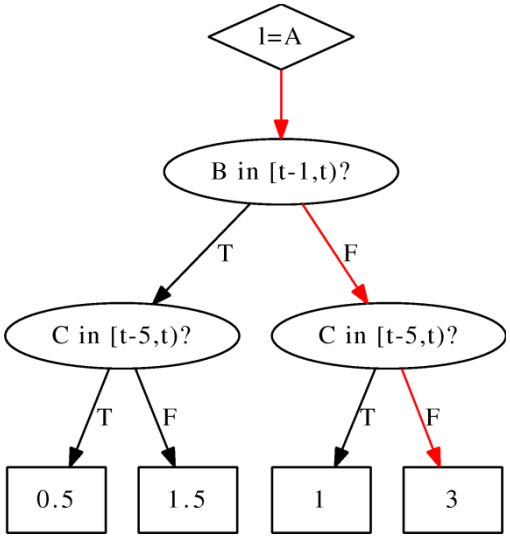


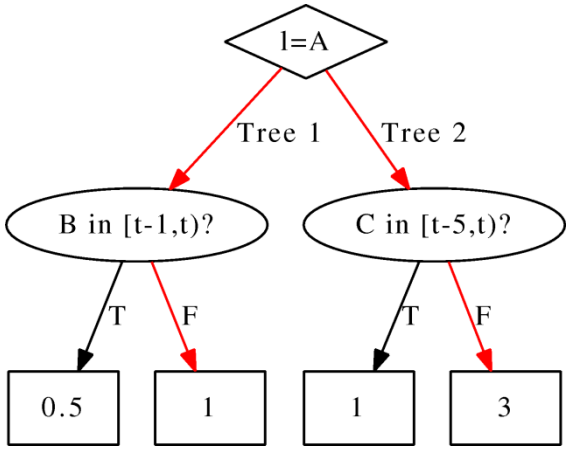
In CTBNs, multiplicative forests (Weiss et al, NIPS 2012):

- Efficiently represent **complex dependencies**
- Empirically **require less data** to learn
- Are learned by **maximizing change in log likelihood**
- Are learned neither in series or in parallel

Multiplicative forests

Represent dependencies with ~~trees~~ **forests**





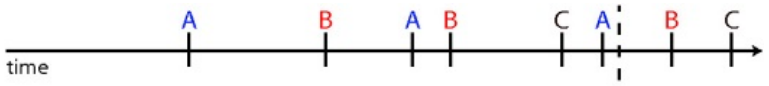
We can apply multiplicative forests to point processes

In CTBNs, multiplicative forests (Weiss et al, NIPS 2012):

- Efficiently represent **complex dependencies**
- Empirically **require less data** to learn
- Are learned by **maximizing change in log likelihood**
- Are learned neither in series or in parallel

Forecasting

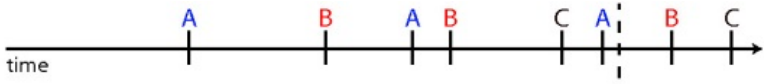
Ex ante forecasting



No labels in the forecast region

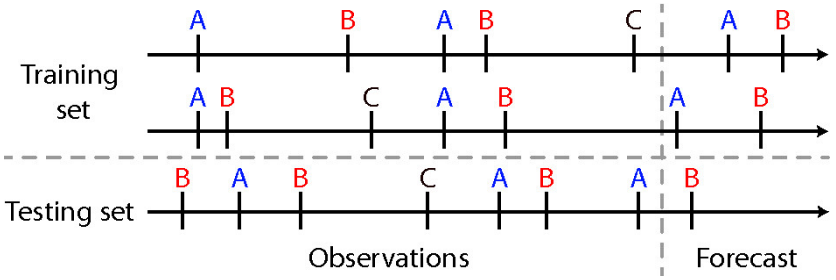
Forecasting

Ex ante forecasting



No labels in the forecast region

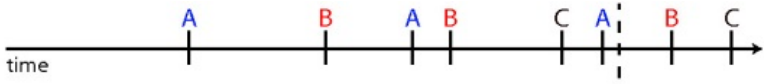
Supervised forecasting



Labels in the forecast region

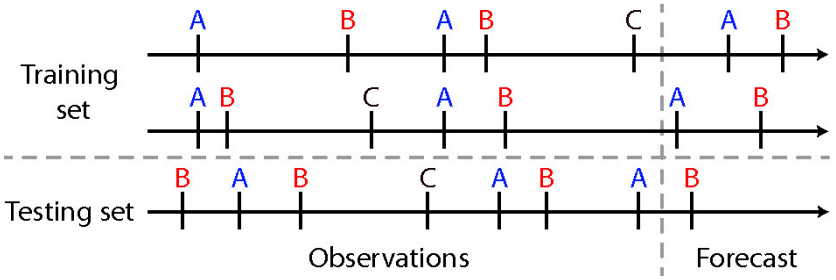
Forecasting

Ex ante forecasting



No labels in the forecast region

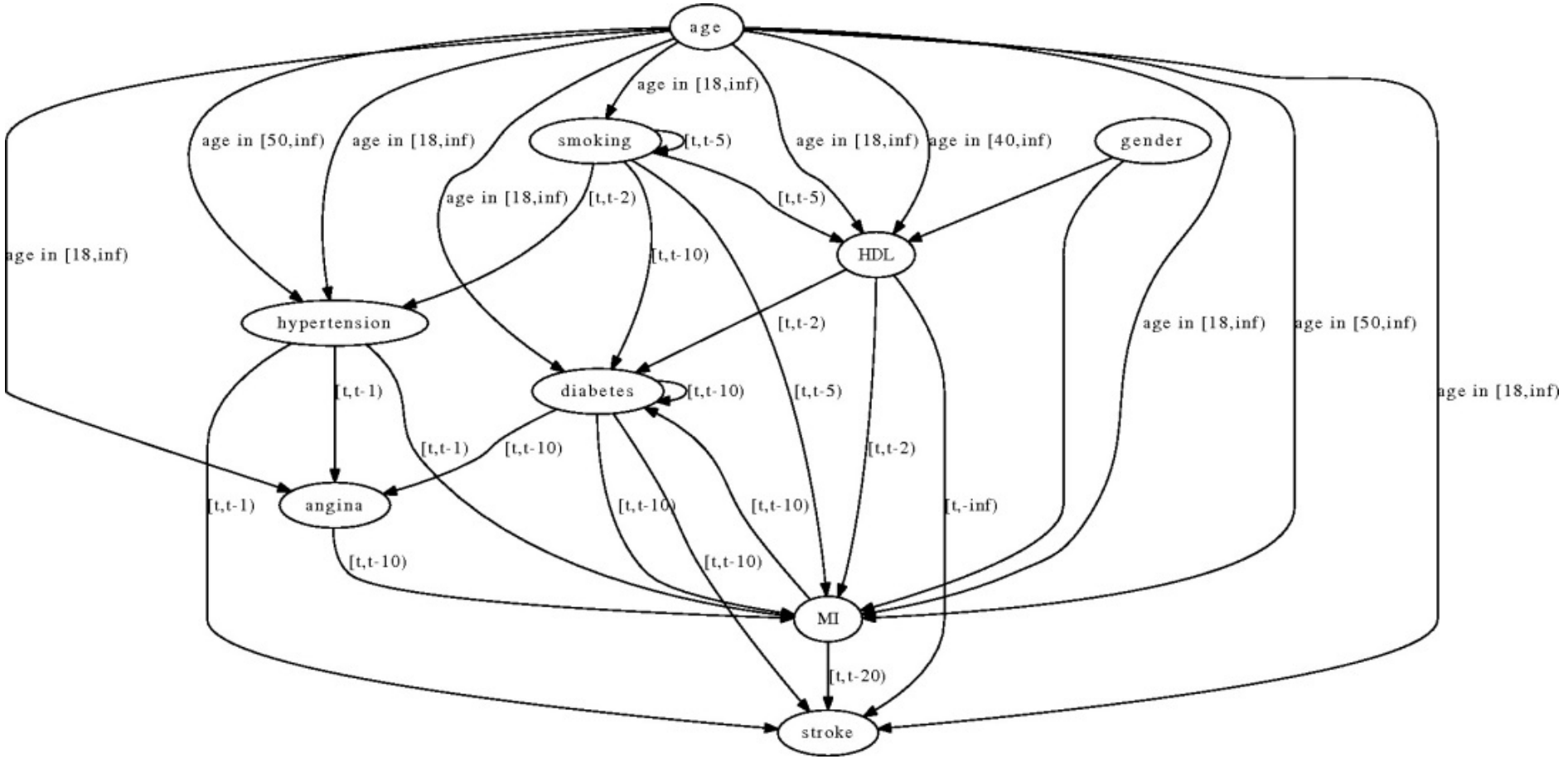
Supervised forecasting



Labels in the forecast region

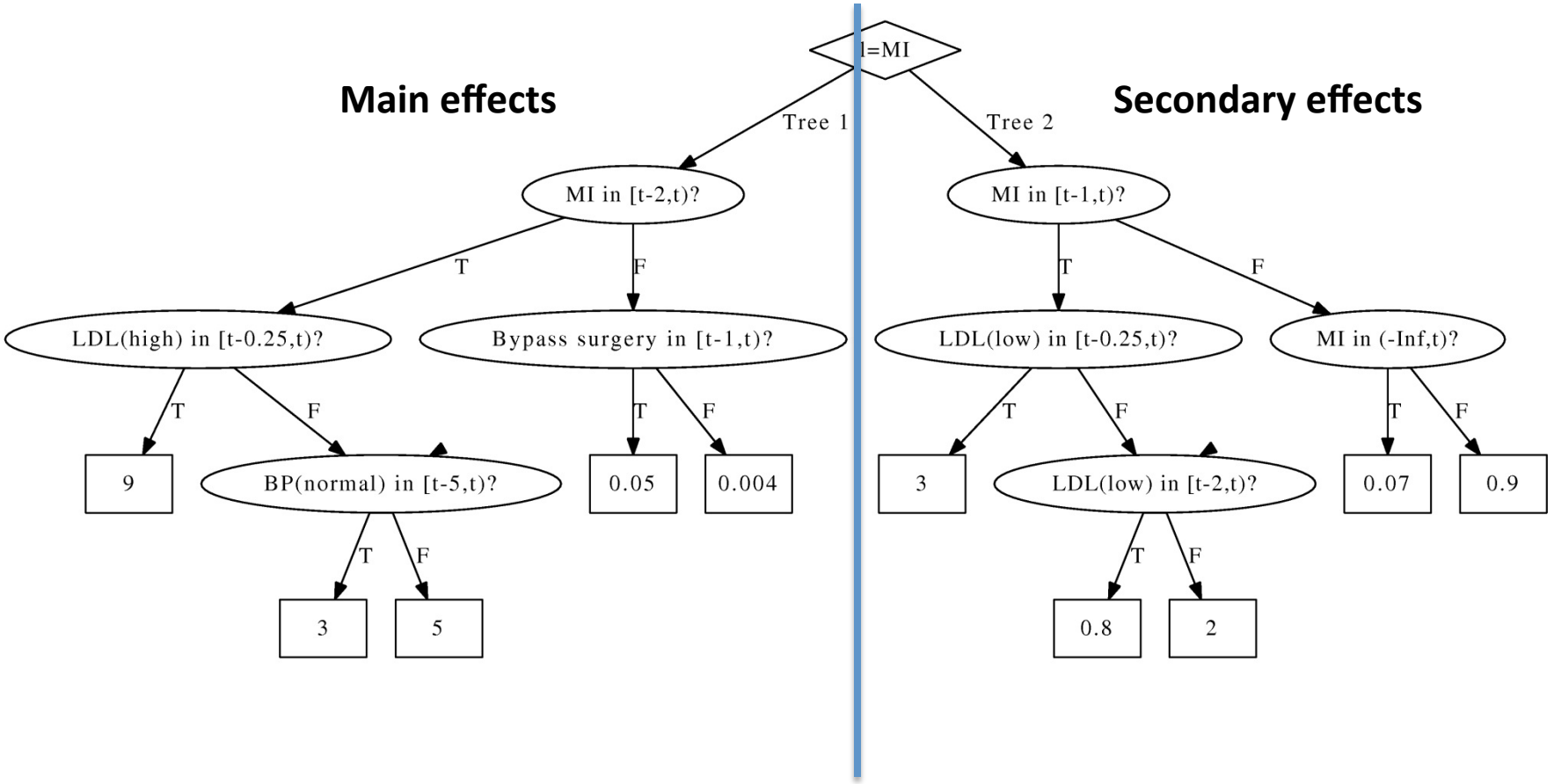
Given data from 1960 to 2005,
will a patient have an MI between 2005 and 2010?

Example CTBN/PCIM Structure



Goal: recover network-dependent event rates – measured by **test set log likelihood**

Result from EHR Data



General Point Process Models (Rasmussen, 2011)

Proposition 2.2 *A conditional intensity function $\lambda^*(t)$ uniquely defines a point process if it satisfies the following conditions for all t and all possible point patterns before t:*

1. *$\lambda^*(t)$ is well-defined and non-negative,*
2. *the integral $\int_{t_n}^{t} \lambda^*(s)\mathrm{d}s$ is well-defined,*
3. *$\int_{t_n}^{t} \lambda^*(s)\mathrm{d}s \to \infty$ for $t \to \infty$.*

Concrete Example

Example 2.3 (Hawkes process) Define a point process by the conditional intensity function

$$\lambda^*(t) = \mu + \alpha \sum_{t_i<t} \exp(-(t - t_i)), \tag{2}$$

where μ and α are positive parameters. Note that each time a new point arrives in this process, the conditional intensity grows by α and then decreases exponentially

Hawkes Process

$$\lambda^*(t) = \mu(t) + \alpha \sum_{t_i<t} \gamma(t - t_i; \beta),$$

where $\mu(t) \geq 0$, $\alpha > 0$, and $\gamma(t; \beta)$ is a density on $(0, \infty)$ depending on some parameter β. For more on the Hawkes process, see e.g. Hawkes (1971b,a, 1972); Hawkes and Oakes (1974).

Hawkes Process

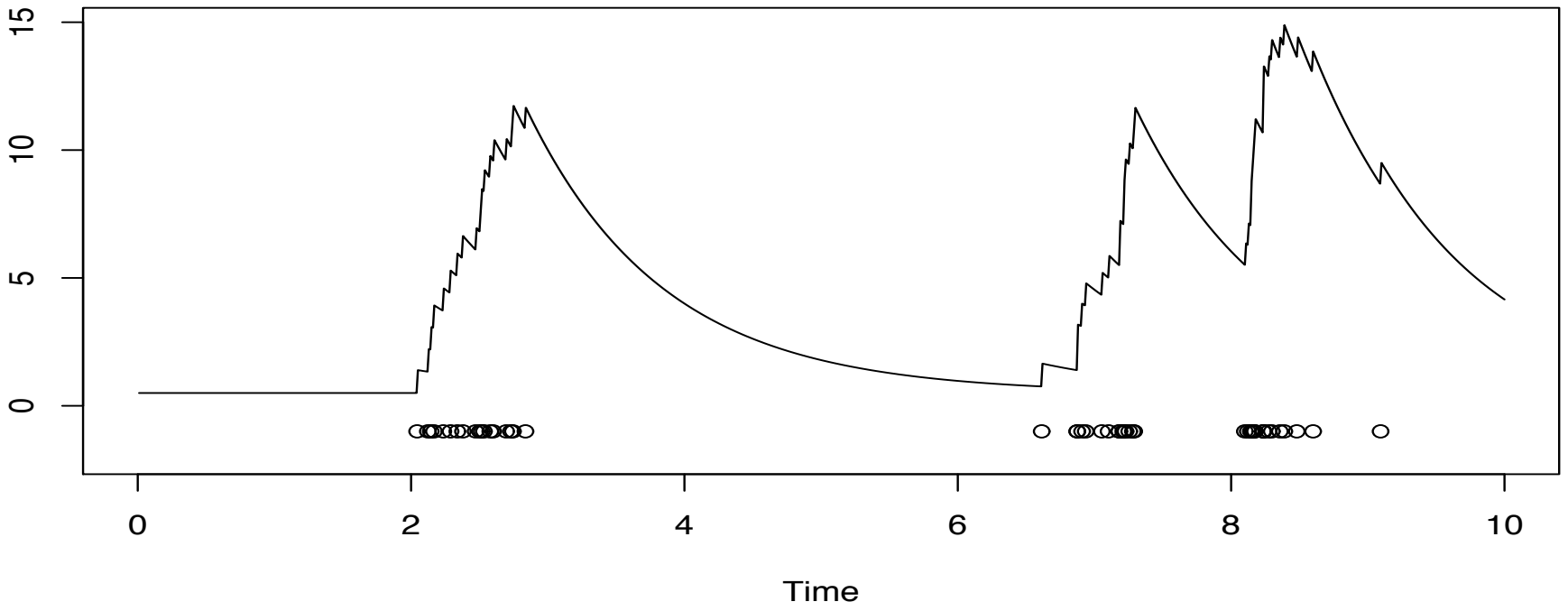


Figure 2: A simulation of the Hawkes process is shown at the bottom of this plot, and the corresponding conditional intensity function is shown in the top. Note that the point pattern is clustered.

One Application

- Let events have more than one type: drugs and diagnoses
- Let **β** be vector of coefficients, one per event type
- Lasso-penalized Poisson regression is similar to lasso-penalized logistic regression and can estimate coefficients to yield λ based on time since prior events of various types
- Being used now for identifying adverse drug events by finding drugs with high coefficient for certain conditions in future (MSCCS, Simpson et al., 2013; current work with Bao, Kuang, and Willett)

Comments on Temporal Models

- DBNs: regular time intervals
- Continuous-time models: inference is hard because fill in values of all variables over all time based on just point evidence
- Point process models: distribution over events (observations) rather than over values of all variables over all time
- Inference and learning are hard: much active research on these and on alternative intensity functions based on history