Graphical Models

Lecture 5:

Template-Based Representations

Andrew McCallum
mccallum@cs.umass.edu

Thanks to Noah Smith and Carlos Guestrin for some slide materials.

Administration

- Homework #3 won't go out until early March. Push back HW#2 due date?
- Lagrange Multipliers?
- Calendar.

BN with Repeated Structure

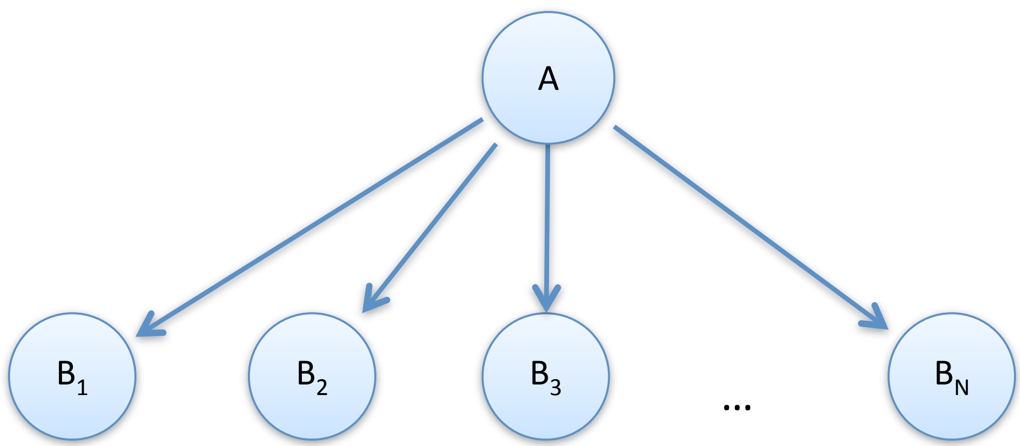
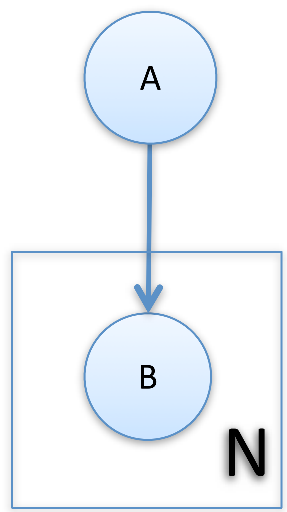
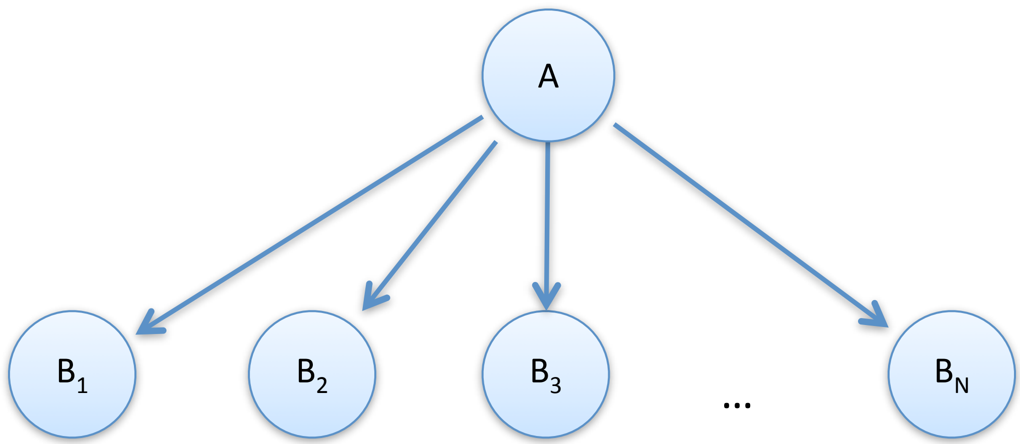


Plate Model

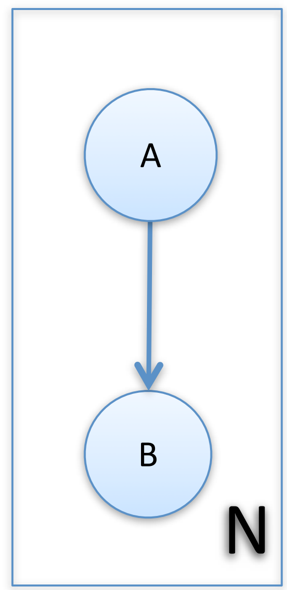


“Unrolled” Ground Network



Ground network

Students and their Grades

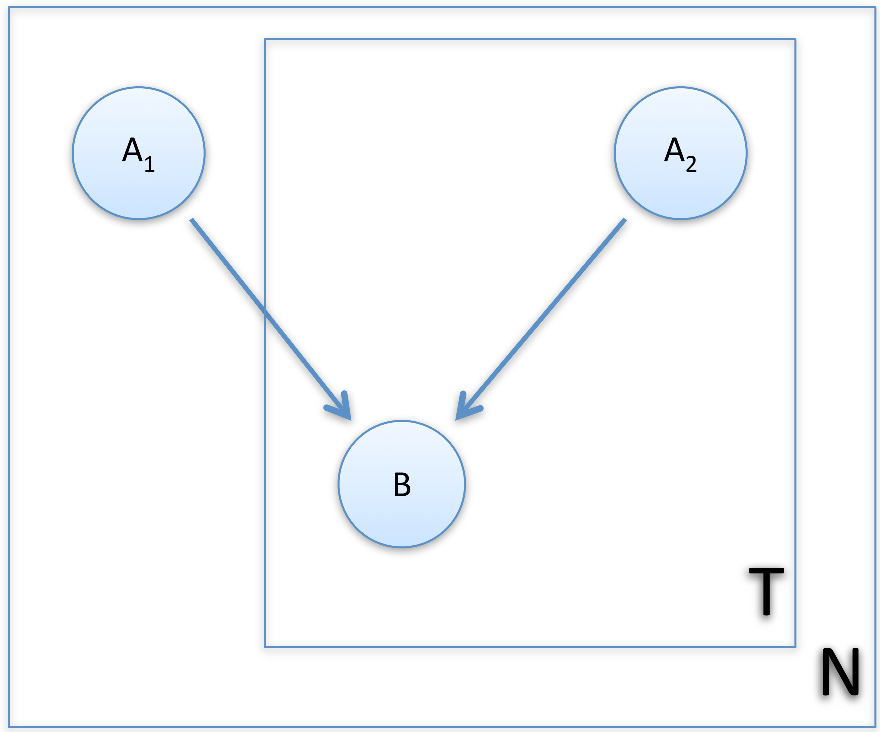


Example: A = student, B = grade

Student, Course, Grade, Difficulty

Each student takes only one course

Nesting

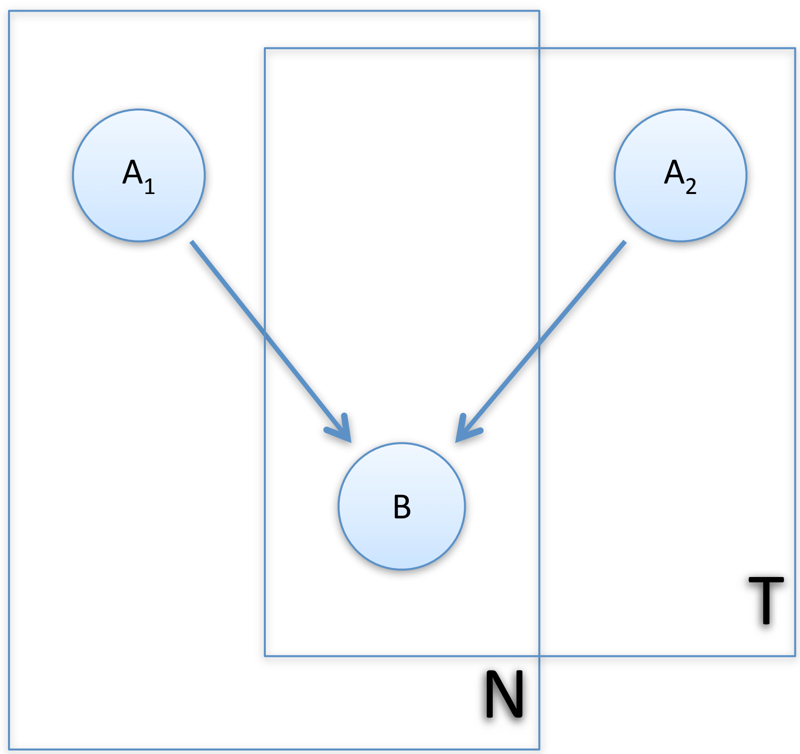


Example: A_1 = course difficulty, A_2 = student aptitude for the area, B = grade

Student, Course, Grade, Difficulty

Multiple courses per student

Intersecting



Example: A_1 = assignment difficulty, A_2 = intelligence, B = grade

Plate Models: Limitations and Alternatives

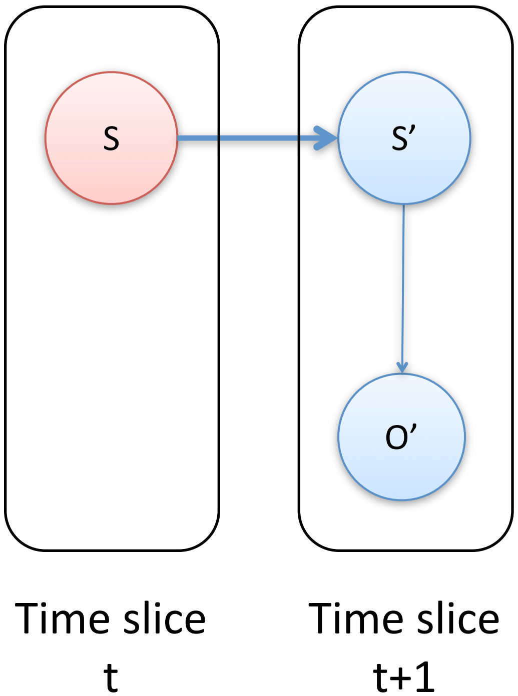
- Limitations:
  - can't have edges between two "copies" of the same variable, (e.g. *position* a time t depends on *position* at time time t-1)
  - can't have edges between particular pairs selected by some other relation, (e.g. Genotype(U_1) depends on Genotype(U_2), where U_2 is mother of U_1.

- Alternatives
  - Dynamic Bayesian Networks (DBNs)
    - Specific to repetitions over time
  - Probabilistic relational models
    - More flexible; see K&F 6.4.2.

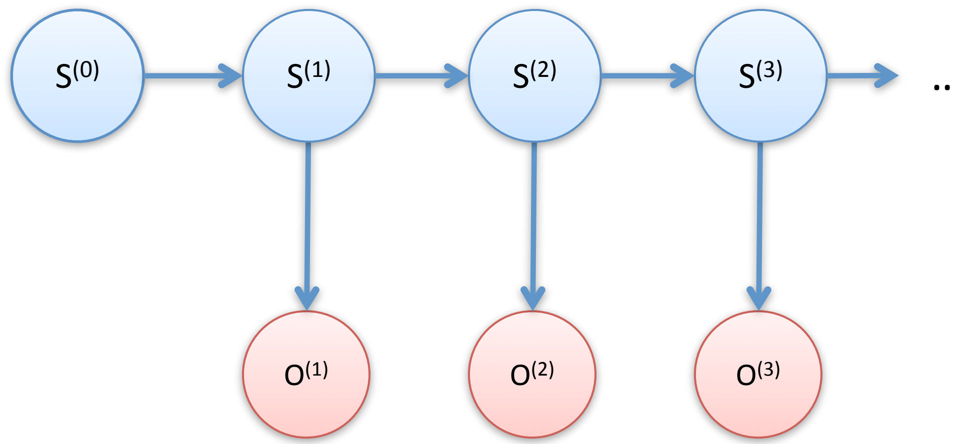
Temporal Models

- **X** takes different values at each (discrete) time step.
  - $\mathbf{X}^{(t)}$ is the random variable at time t
- Markov Assumption:
  $\mathbf{X}^{(t+1)} \perp \{\mathbf{X}^{(0)}, \ldots, \mathbf{X}^{(t-1)}\} \mid \mathbf{X}^{(t)}$
- Stationary Assumption (aka *time invariant* or *homogeneous*)
  $P(\mathbf{X}^{(t+1)} \mid \mathbf{X}^{(t)})$ is the same for all t.
- Can use *conditional Bayesian network* to define
  $P(\mathbf{X}^{(t+1)} \mid \mathbf{X}^{(t)})$

Hidden Markov Model



2-time-slice *conditional BN*

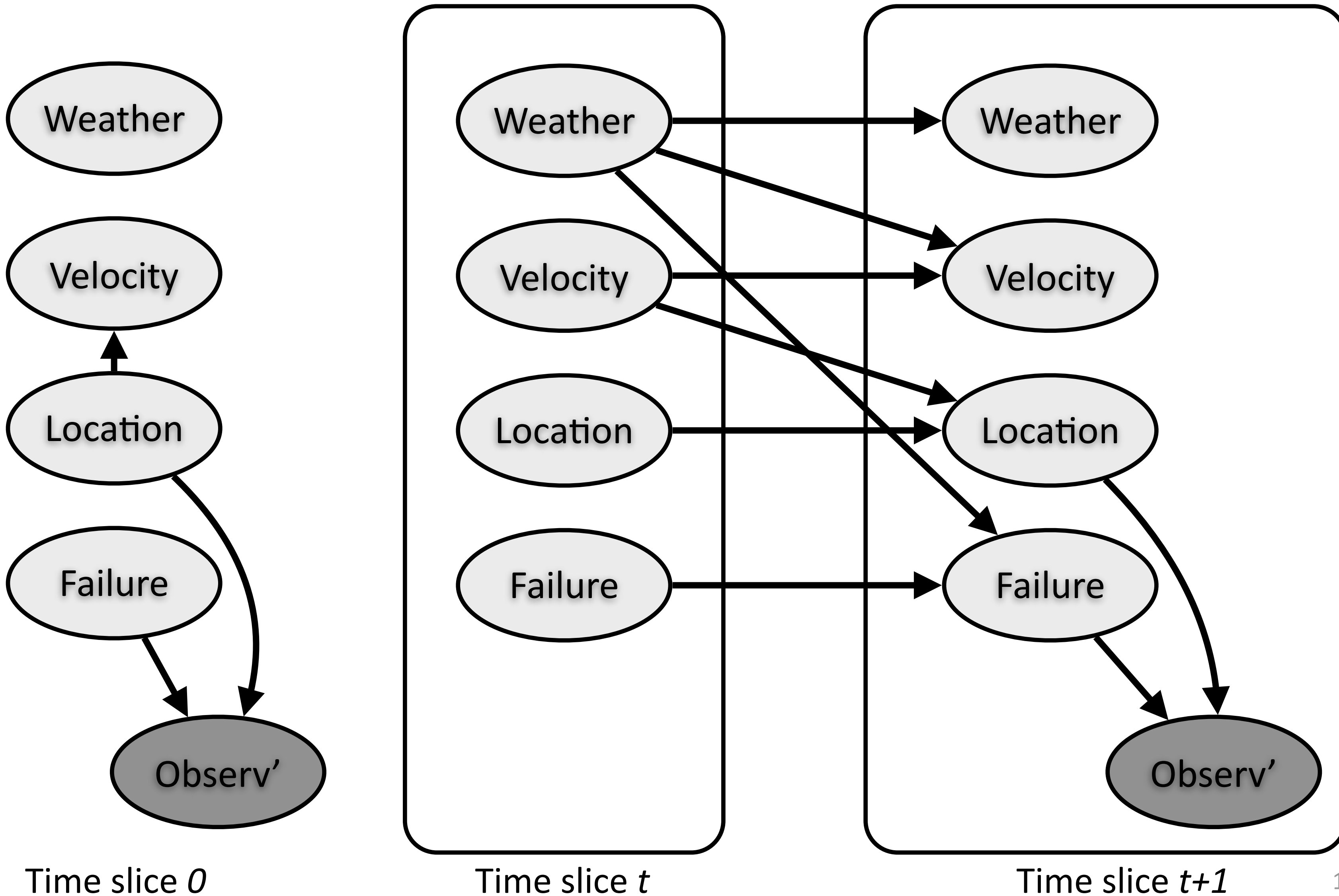


unrolled or ground Bayesian network

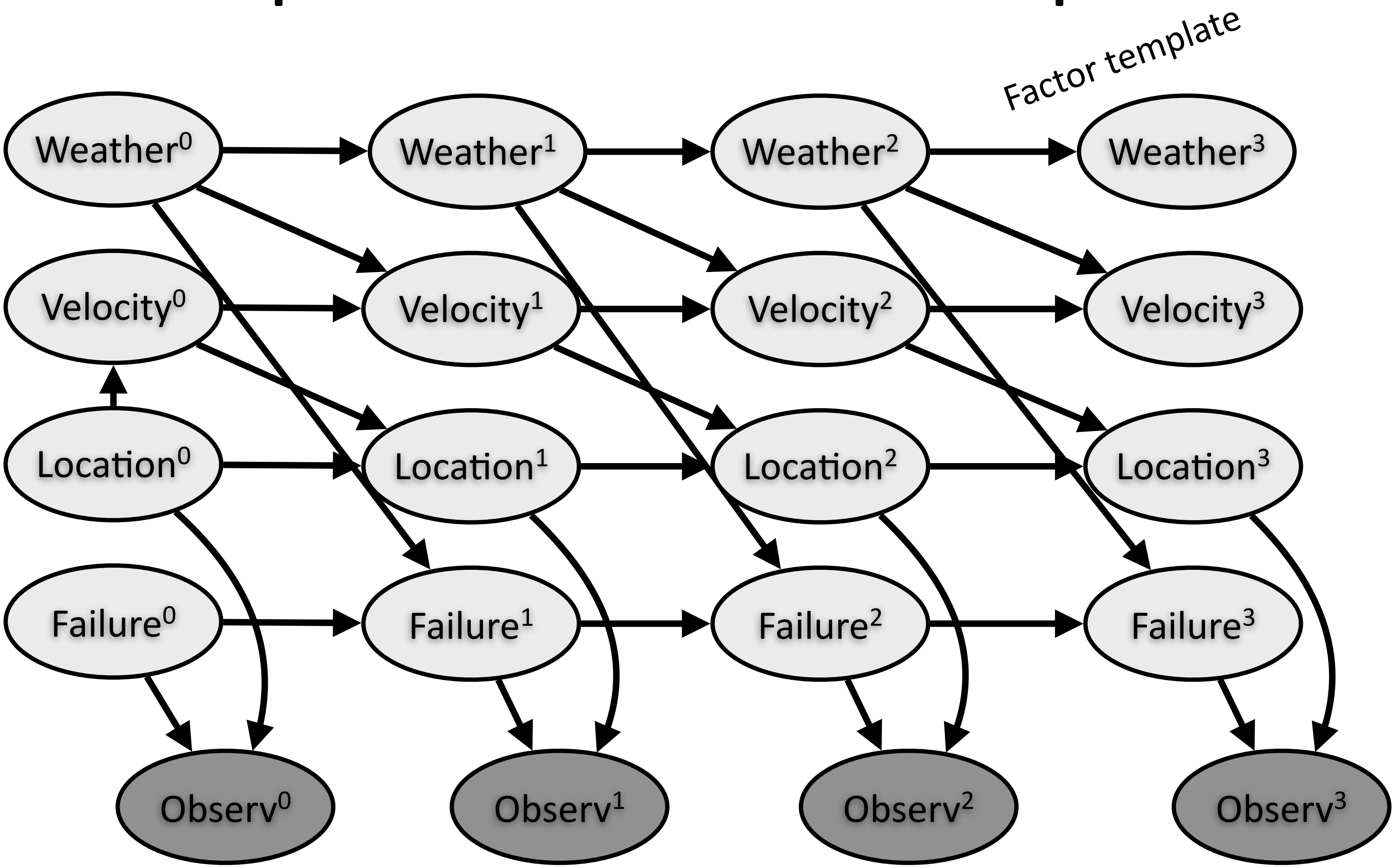
Dynamic Bayesian Network

- Bayesian network over $\mathbf{X}^{(0)}$,
  conditional Bayesian network for $\mathbf{X}^{(t+1)}$ given $X^{(t)}$
  (2-time-slice)
  - HMM is a special case.
  - Kalman filter (linear dynamical system) is a special case.

Example: DBN for vehicle position

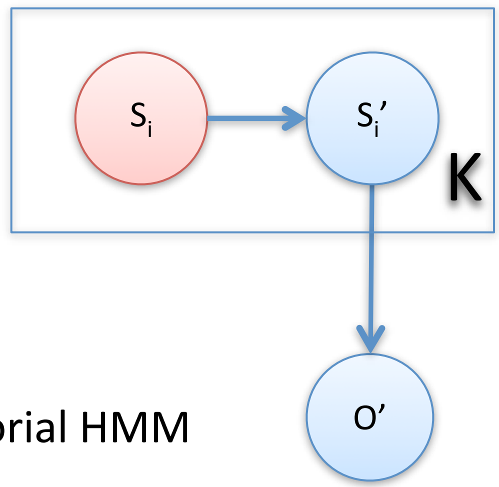


Example: DBN for vehicle position

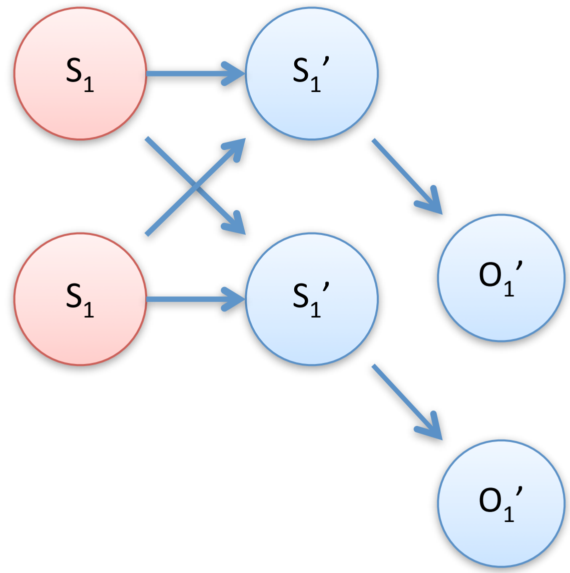


Unrolled over 3 steps

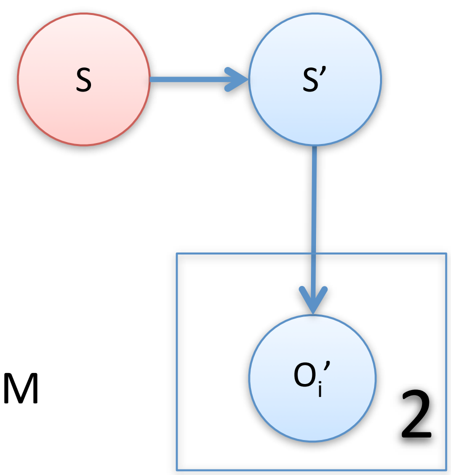
Dynamic Bayesian Networks



Factorial HMM



Coupled HMM



Pair HMM

Probabilistic Relational Models

- Contingent Dependency
  - specifies the context in which some dependency holds, with a “guard”—a formula that must hold for the dependency to be applicable.
  - e.g. Location(V) depends on Location(U) contingent on Precedes(U,V)
  - e.g. Genotype(V) depends on Genotype(U) contingent on Mother(U,V)
- Relational Uncertainty (one kind of *structural uncertainty*)
  - The “guard” predicates are random variables!

Object Uncertainty

[Milch et al “BLOG”]

- The set of objects is not predetermined.
  - Get list of authors in 100 BibTeX files.
    “Stuart Russell” “Stuart Rusell” “S. Russell”
    How many people are mentioned?
- Introduce
  person-objects (represents entity)
  person-reference objects (represents mention)
  refers-to(m,o) relation
- Model generates (a) # of people, (b) person objects, (c) their reference objects.

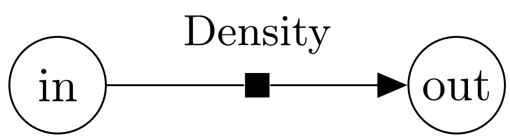
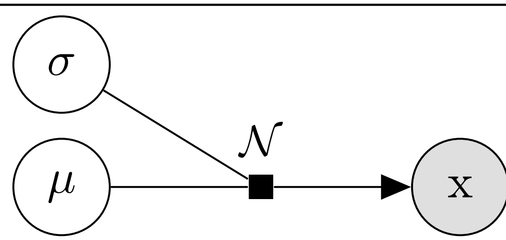
Directed Factor Graph Notation

[Laura Dietz 2010]

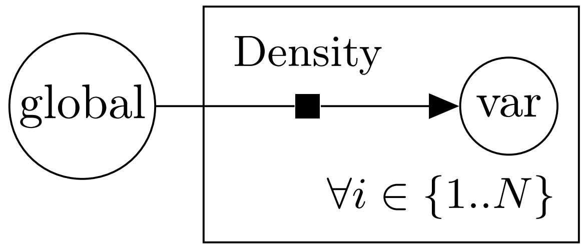
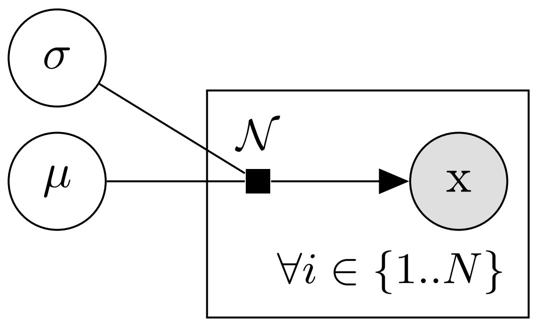
Variables and Constants

	Directed factor graph	Pseudocode
Llatent variable / latent parameter	var	
Observed variable	obs	
Constant / hyper parameter	const	

Factors and Densities

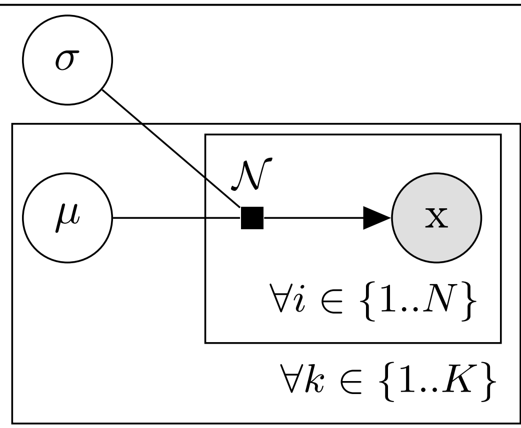
	Directed factor graph	Pseudocode
Factor with one input parameter		**draw** out $\sim$ Density(in)
Example: Gaussian		**draw** $x \sim \mathcal{N}(\mu, \sigma)$

Replication with Plates

	Directed factor graph	Pseudocode
Plate		**for all** $\forall i \in \{1..N\}$ **do** **draw** $\mathrm{var}_i \sim \mathrm{global}$
Example: repeated Gaussian		**for all** $\forall i \in \{1..N\}$ **do** **draw** $x_i \sim \mathcal{N}(\mu, \sigma)$

Nested Plates

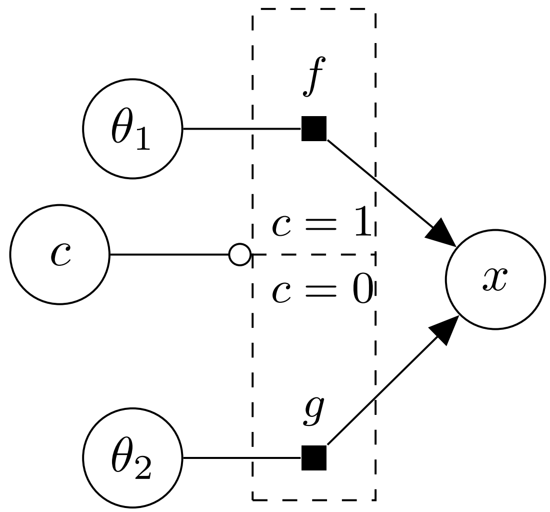
	Directed factor graph	Pseudocode

Nested plates		**for all** $\forall k \in \{1..K\}$ **do**
		    for all $\forall i \in \{1..N\}$ **do**
		        draw $x_{k,i} \sim \mathcal{N}(\mu_k, \sigma)$

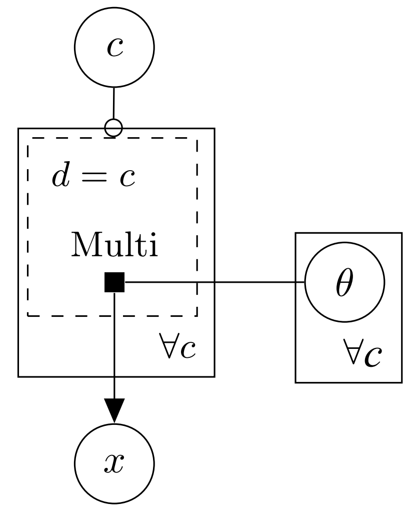
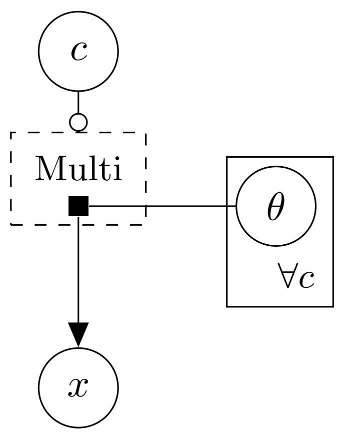
Conditioning with *Gates*

Minka & Winn 2008

	Directed factor graph	Pseudocode

Unrolled boolean gate		**if** $c = 1$ **then**
		    draw $x \sim f(\theta_1)$
		else
		    draw $x \sim g(\theta_2)$

Plates & Gates (and implicit combo)

	Directed factor graph	Pseudocode
Replicated gate		1: **draw** $x \sim \mathrm{Multi}(\theta_c)$
Implicit notation for replicating gates		1: **draw** $x \sim \mathrm{Multi}(\theta_c)$

Latent Dirichlet Allocation

[Blei, Ng, Jordan]