Hierarchical Networks

Marco Cattaneo
Department of Statistics, LMU Munich
cattaneo@stat.uni-muenchen.de

September 10, 2009

hierarchical model

- two levels:

  probabilistic level: set $\mathcal{P}$ of probability measures P on $(\Omega, \mathcal{A})$

  possibilistic level: likelihood function $lik : \mathcal{P} \to (0\,,\infty)$

hierarchical model

- two levels:

  probabilistic level: set $\mathcal{P}$ of probability measures P on $(\Omega, \mathcal{A})$

  possibilistic level: likelihood function $lik : \mathcal{P} \to (0\,,\infty)$

- likelihood function $=:$ membership function of fuzzy set of probability measures

hierarchical model

- two levels:

  probabilistic level: set $\mathcal{P}$ of probability measures P on $(\Omega, \mathcal{A})$

  possibilistic level: likelihood function $lik : \mathcal{P} \to (0\,,\infty)$

- likelihood function $=:$ membership function of fuzzy set of probability measures

- inference about the value $g(P)$ of $g : \mathcal{P} \to \mathcal{G}$ (**extension principle**):

  fuzzy subset of $\mathcal{G}$ with as membership function the **profile** likelihood function $lik_g : \mathcal{G} \to (0\,,\infty)$ defined by

$$lik_g(\gamma) \propto \sup_{P \in \mathcal{P}\,:\, g(P) = \gamma} lik(P)$$

hierarchical model

- two levels:

  probabilistic level: set $\mathcal{P}$ of probability measures P on $(\Omega, \mathcal{A})$

  possibilistic level: likelihood function $lik : \mathcal{P} \to (0\,,\infty)$

- likelihood function =: membership function of fuzzy set of probability measures

- inference about the value $g(P)$ of $g : \mathcal{P} \to \mathcal{G}$ (**extension principle**):

  fuzzy subset of $\mathcal{G}$ with as membership function the **profile** likelihood function $lik_g : \mathcal{G} \to (0\,,\infty)$ defined by
  $$lik_g(\gamma) \propto \sup_{P \in \mathcal{P}\,:\,g(P)=\gamma} lik(P)$$

- interval probability model: special case with *lik* constant

hierarchical network

- directed acyclic graph with nodes $X_1 \in \mathcal{X}_1, \ldots, X_k \in \mathcal{X}_k$, such that to each node X_i are associated a set $\mathcal{P}_i$ of stochastic kernels P_i from $\mathcal{PA}_i$ (the set of all possible values of the parents of X_i) to $\mathcal{X}_i$, and a likelihood function $lik_i : \mathcal{P}_i \to (0\,,\infty)$

hierarchical network

- directed acyclic graph with nodes $X_1 \in \mathcal{X}_1, \ldots, X_k \in \mathcal{X}_k$, such that to each node X_i are associated a set $\mathcal{P}_i$ of stochastic kernels P_i from $\mathcal{PA}_i$ (the set of all possible values of the parents of X_i) to $\mathcal{X}_i$, and a likelihood function $lik_i : \mathcal{P}_i \to (0\,,\infty)$

- resulting **hierarchical model**:

probabilistic level: $\mathcal{P} = \{P_{P_1,\ldots,P_k} : P_1 \in \mathcal{P}_1, \ldots, P_k \in \mathcal{P}_k\}$ with

$$P_{P_1,\ldots,P_k}(x_1, \ldots, x_k) = \prod_{i=1}^{k} P_i\,(x_i \mid pa_i(x_1, \ldots, x_k))$$

possibilistic level: $lik : \mathcal{P} \to (0\,,\infty)$ with

$$lik(P) \propto \sup_{\substack{P_1 \in \mathcal{P}_1, \ldots, P_k \in \mathcal{P}_k : \\ P_{P_1,\ldots,P_k} = P}} \prod_{i=1}^{k} lik_i(P_i)$$

hierarchical network

- directed acyclic graph with nodes $X_1 \in \mathcal{X}_1, \ldots, X_k \in \mathcal{X}_k$, such that to each node X_i are associated a set $\mathcal{P}_i$ of stochastic kernels P_i from $\mathcal{PA}_i$ (the set of all possible values of the parents of X_i) to $\mathcal{X}_i$, and a likelihood function $lik_i : \mathcal{P}_i \to (0\,,\infty)$
- resulting **hierarchical model**:

probabilistic level: $\mathcal{P} = \{P_{P_1,\ldots,P_k} : P_1 \in \mathcal{P}_1, \ldots, P_k \in \mathcal{P}_k\}$ with

$$P_{P_1,\ldots,P_k}(x_1,\ldots,x_k) = \prod_{i=1}^{k} P_i\left(x_i \,|\, pa_i(x_1,\ldots,x_k)\right)$$

possibilistic level: $lik : \mathcal{P} \to (0\,,\infty)$ with

$$lik(P) \propto \sup_{\substack{P_1 \in \mathcal{P}_1,\ldots,P_k \in \mathcal{P}_k :\\ P_{P_1,\ldots,P_k} = P}} \prod_{i=1}^{k} lik_i(P_i)$$

- credal network (with strong independence): special case with lik_i constant

updating

- event $A \in \mathcal{A}$ observed:

$$\text{probabilistic level:} \quad \mathcal{P} \ \rightsquigarrow \ \mathcal{P}' = \{P(\cdot \mid A) : P \in \mathcal{P},\ P(A) > 0\}$$

$$\text{possibilistic level:} \quad \mathit{lik} \ \rightsquigarrow \ \mathit{lik}' : \mathcal{P}' \to (0\,,\infty) \text{ with}$$

$$\mathit{lik}'(P') \propto \sup_{P \in \mathcal{P} \,:\, P(\cdot \mid A) = P'} \mathit{lik}(P)\, P(A)$$

updating

- event $A \in \mathcal{A}$ observed:

$$\text{probabilistic level:} \quad \mathcal{P} \ \rightsquigarrow \ \mathcal{P}' = \{P(\cdot \mid A) : P \in \mathcal{P},\ P(A) > 0\}$$

$$\text{possibilistic level:} \quad \mathit{lik} \ \rightsquigarrow \ \mathit{lik}' : \mathcal{P}' \to (0\,,\infty) \text{ with}$$

$$\mathit{lik}'(P') \propto \sup_{P \in \mathcal{P} \,:\, P(\cdot \mid A) = P'} \mathit{lik}(P)\, P(A)$$

- lik constant $\quad \not\Rightarrow \quad$ lik' constant

updating

- event $A \in \mathcal{A}$ observed:

$$\text{probabilistic level:} \quad \mathcal{P} \rightsquigarrow \mathcal{P}' = \{P(\cdot \mid A) : P \in \mathcal{P},\ P(A) > 0\}$$

$$\text{possibilistic level:} \quad lik \rightsquigarrow lik' : \mathcal{P}' \to (0\,,\infty) \text{ with}$$

$$lik'(P') \propto \sup_{P \in \mathcal{P}\,:\, P(\cdot \mid A) = P'} lik(P)\, P(A)$$

- lik constant $\quad \not\Rightarrow \quad lik'$ constant

- **d-separation implies conditional irrelevance in hierarchical networks**

updating

- event $A \in \mathcal{A}$ observed:

$$\text{probabilistic level:} \quad \mathcal{P} \;\rightsquigarrow\; \mathcal{P}' = \{P(\cdot \,|\, A) : P \in \mathcal{P},\; P(A) > 0\}$$

$$\text{possibilistic level:} \quad lik \;\rightsquigarrow\; lik' : \mathcal{P}' \to (0\,,\infty) \text{ with}$$

$$lik'(P') \propto \sup_{P \in \mathcal{P} \,:\, P(\cdot \,|\, A) = P'} lik(P)\, P(A)$$

- *lik* constant $\quad \not\Rightarrow \quad$ *lik'* constant

- **d-separation implies conditional irrelevance in hierarchical networks**

- probabilistic level: d-separation implies conditional irrelevance in credal networks (with strong independence)

alternative description of hierarchical model

- description as **set of measures** (not unique):

  set $\mathcal{M}$ of measures μ on $(\Omega, \mathcal{A})$ with $\mu(\Omega) \in (0, \infty)$

alternative description of hierarchical model

- description as **set of measures** (not unique):

  set $\mathcal{M}$ of measures μ on $(\Omega, \mathcal{A})$ with $\mu(\Omega) \in (0, \infty)$

- two levels $\rightsquigarrow$ set of measures:

$$\mathcal{M} = \{ lik(P)\, P : P \in \mathcal{P} \}$$

alternative description of hierarchical model

- description as **set of measures** (not unique):

$$\text{set } \mathcal{M} \text{ of measures } \mu \text{ on } (\Omega, \mathcal{A}) \text{ with } \mu(\Omega) \in (0, \infty)$$

- two levels $\rightsquigarrow$ set of measures:

$$\mathcal{M} = \{ \mathit{lik}(P)\, P : P \in \mathcal{P} \}$$

- set of measures $\rightsquigarrow$ two levels:

$$\begin{aligned} \text{probabilistic level:} \quad & \mathcal{P} = \left\{ \frac{\mu}{\mu(\Omega)} : \mu \in \mathcal{M} \right\} \\ \text{possibilistic level:} \quad & \mathit{lik} : \mathcal{P} \to (0\,,\infty) \text{ with} \\ & \mathit{lik}(P) \propto \sup_{\mu \in \mathcal{M}\,:\, \frac{\mu}{\mu(\Omega)} = P} \mu(\Omega) \end{aligned}$$

alternative description of hierarchical model

- description as **set of measures** (not unique):

  set $\mathcal{M}$ of measures μ on $(\Omega, \mathcal{A})$ with $\mu(\Omega) \in (0, \infty)$

- two levels $\rightsquigarrow$ set of measures:

$$\mathcal{M} = \{ \mathit{lik}(P)\, P : P \in \mathcal{P} \}$$

- set of measures $\rightsquigarrow$ two levels:

$$\text{probabilistic level:} \quad \mathcal{P} = \left\{ \frac{\mu}{\mu(\Omega)} : \mu \in \mathcal{M} \right\}$$

$$\text{possibilistic level:} \quad \mathit{lik} : \mathcal{P} \to (0\,, \infty) \text{ with}$$

$$\mathit{lik}(P) \propto \sup_{\mu \in \mathcal{M} \,:\, \frac{\mu}{\mu(\Omega)} = P} \mu(\Omega)$$

- interval probability model: special case with normalized measures only

convex polytope of measures

- **updating** of set of measures (event $A \in \mathcal{A}$ observed):

$$\mathcal{M} \;\rightsquigarrow\; \mathcal{M}' = \{\mu(\cdot \cap A) : \mu \in \mathcal{M},\, \mu(A) > 0\}$$

convex polytope of measures

- **updating** of set of measures (event $A \in \mathcal{A}$ observed):

$$\mathcal{M} \ \rightsquigarrow \ \mathcal{M}' = \{\mu(\cdot \cap A) : \mu \in \mathcal{M},\, \mu(A) > 0\}$$

- when $\mathcal{M} = \mathrm{ch}\left(\{\mu_i : i \in \{1, \ldots, m\}\}\right)$ is a convex polytope, it suffices to update the **extreme points**:

$$\mathcal{M}' = \mathrm{ch}\left(\{\mu_i(\cdot \cap A) : i \in \{1, \ldots, m\}\,,\ \mu_i(A) > 0\}\right)$$

convex polytope of measures

- **updating** of set of measures (event $A \in \mathcal{A}$ observed):

$$\mathcal{M} \ \rightsquigarrow \ \mathcal{M}' = \{\mu(\cdot \cap A) : \mu \in \mathcal{M},\, \mu(A) > 0\}$$

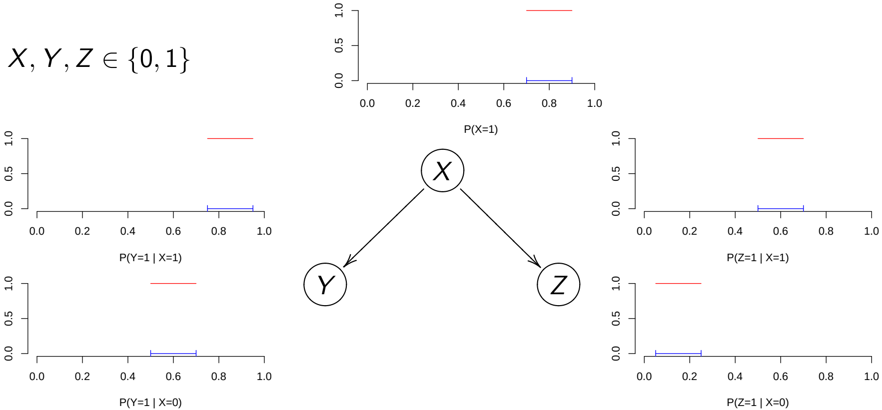
- when $\mathcal{M} = \mathrm{ch}\left(\{\mu_i : i \in \{1, \ldots, m\}\}\right)$ is a convex polytope, it suffices to update the **extreme points**:

$$\mathcal{M}' = \mathrm{ch}\left(\{\mu_i(\cdot \cap A) : i \in \{1, \ldots, m\}\,,\, \mu_i(A) > 0\}\right)$$

- when $\mathcal{M} = \mathrm{ch}(\{lik(P_i)\, P_i : i \in \{1, \ldots, m\}\})$ is a convex polytope, the profile likelihood function for the expected value $E_P(X)$ of a random variable X is **piecewise hyperbolic** and determined by the pairs $(E_{P_i}(X),\, lik(P_i))$ with $i \in \{1, \ldots, m\}$

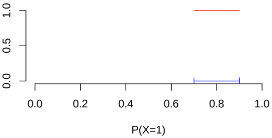
example with convex polytope of measures

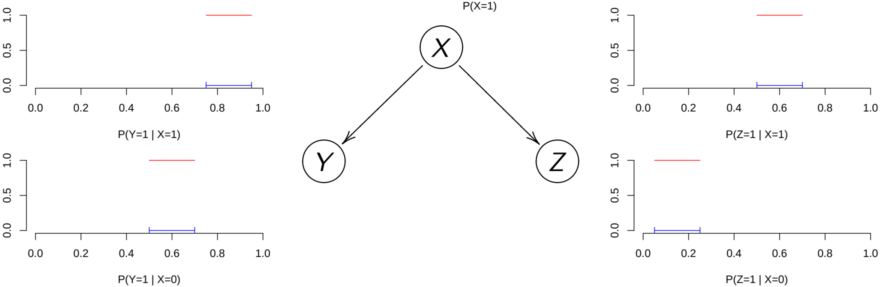
$X, Y, Z \in \{0, 1\}$



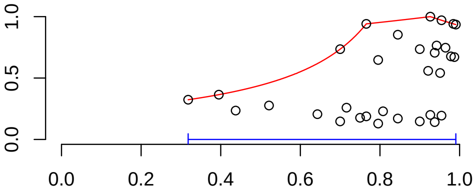
example with convex polytope of measures

$X, Y, Z \in \{0, 1\}$

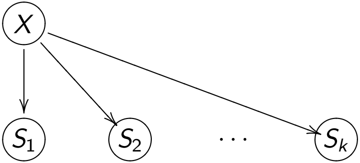




$\Rightarrow \ P(X = 1 \mid Y = 0,\ Z = 1):$

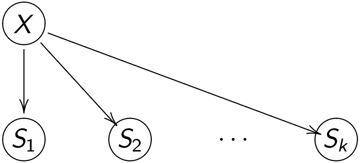


sensors example (Antonucci et al., 2007)



$X, S_1, \dots, S_k \in \{0, 1\}$
$P(X = 1) = \frac{1}{2}$
$P(S_i = x \mid X = x) \geq 0.9$
for all $x \in \{0, 1\}$, $i \in \{1, \dots, k\}$

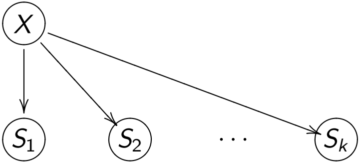
sensors example (Antonucci et al., 2007)



$X, S_1, \dots, S_k \in \{0, 1\}$
$P(X = 1) = \frac{1}{2}$
$P(S_i = x \mid X = x) \geq 0.9$
for all $x \in \{0, 1\}$, $i \in \{1, \dots, k\}$

- **interval probability** updating:
  $P(X = 1 \mid S_1 = s_1, \dots, S_k = s_k)$ is almost 1 if $s_1 = \dots = s_k = 1$, almost 0 if $s_1 = \dots = s_k = 0$, and vacuous otherwise

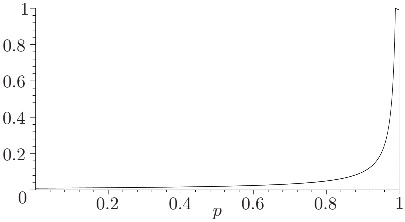
sensors example (Antonucci et al., 2007)



$X, S_1, \ldots, S_k \in \{0, 1\}$
$P(X = 1) = \frac{1}{2}$
$P(S_i = x \mid X = x) \geq 0.9$
for all $x \in \{0, 1\}$, $i \in \{1, \ldots, k\}$

- **interval probability** updating:
  $P(X = 1 \mid S_1 = s_1, \ldots, S_k = s_k)$ is almost 1 if $s_1 = \ldots = s_k = 1$, almost 0 if $s_1 = \ldots = s_k = 0$, and vacuous otherwise

- **hierarchical** updating: e.g. $P(X = 1 \mid S_1 = S_2 = S_3 = 1,\ S_4 = 0)$:



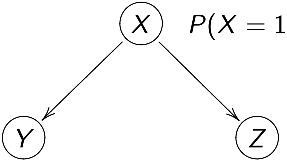
example with training data

X	Y	Z	#
0	0	0	21
0	0	1	6
0	1	0	30
0	1	1	7
1	0	0	9
1	0	1	15
1	1	0	5
1	1	1	7
			100

example with training data

X	Y	Z	#
0	0	0	21
0	0	1	6
0	1	0	30
0	1	1	7
1	0	0	9
1	0	1	15
1	1	0	5
1	1	1	7
			100

simulated according to:

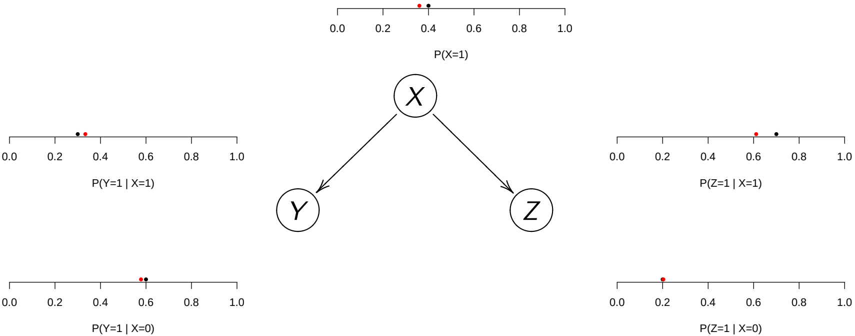


$$P(X=1)=0.4$$

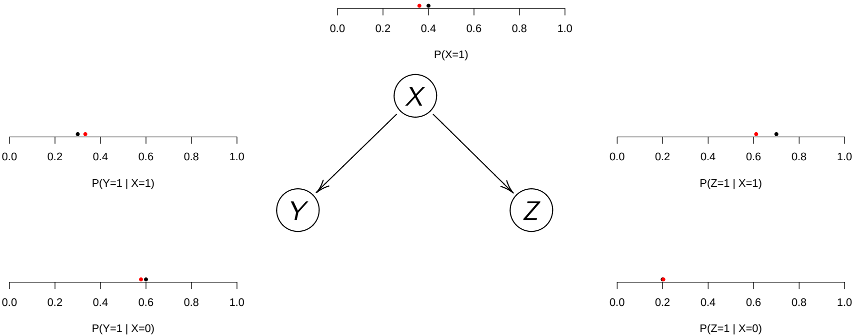
$$P(Y=1\,|\,X=1)=0.3$$
$$P(Y=1\,|\,X=0)=0.6$$

$$P(Z=1\,|\,X=1)=0.7$$
$$P(Z=1\,|\,X=0)=0.2$$

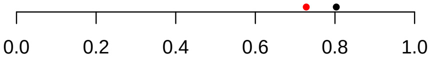
Bayesian network via MLE



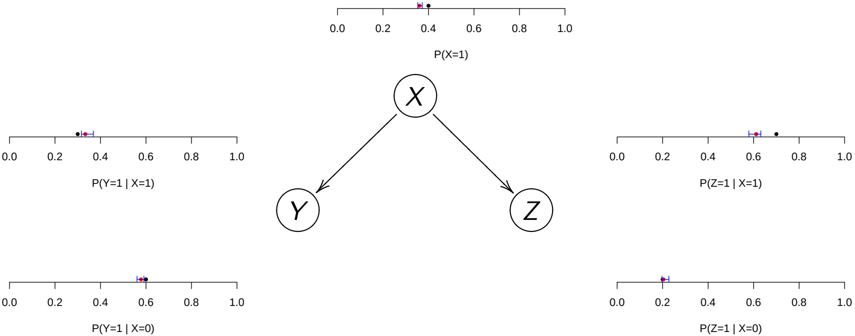
Bayesian network via MLE



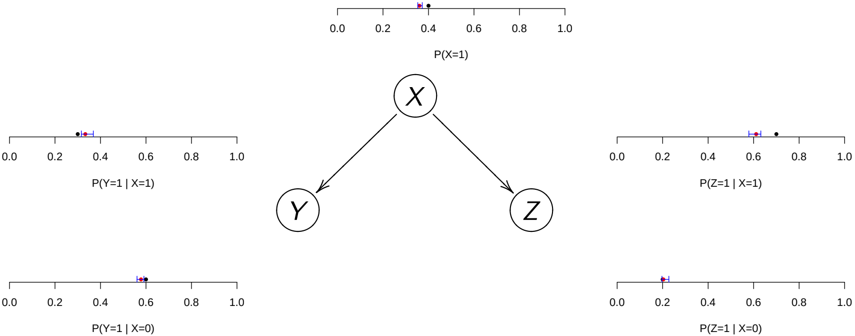
$\Rightarrow \; P(X = 1 \mid Y = 0, \, Z = 1)$:



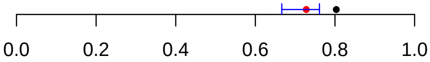
credal network via IDM (with $s = 2$)



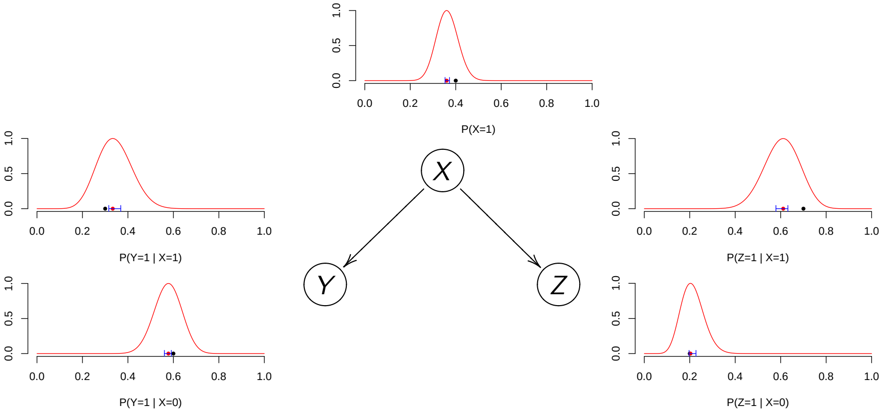
credal network via IDM (with $s = 2$)



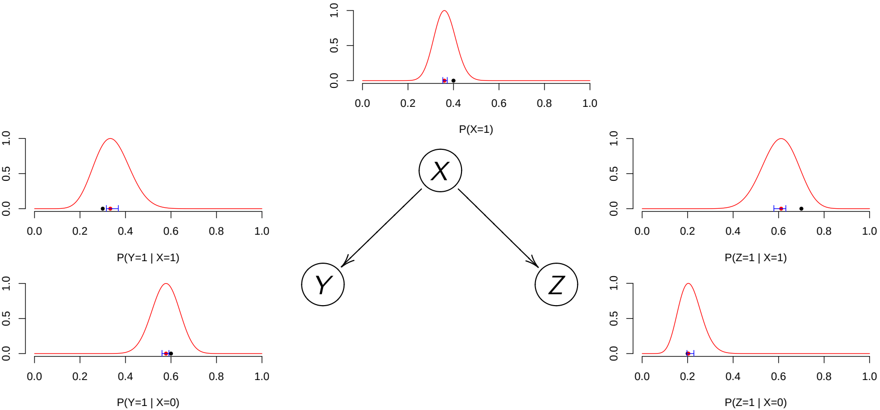
$\Rightarrow \; P(X = 1 \mid Y = 0, \, Z = 1):$



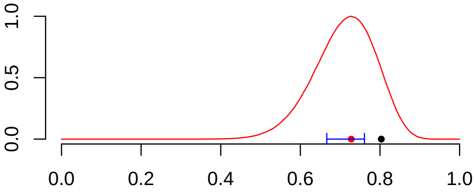
hierarchical network



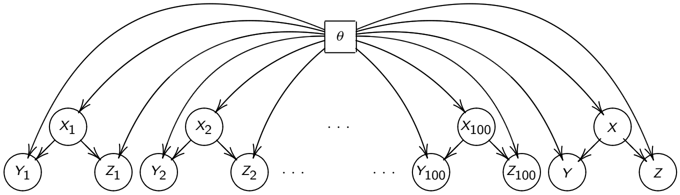
hierarchical network



$$\Rightarrow \; P(X=1 \mid Y=0,\, Z=1):$$

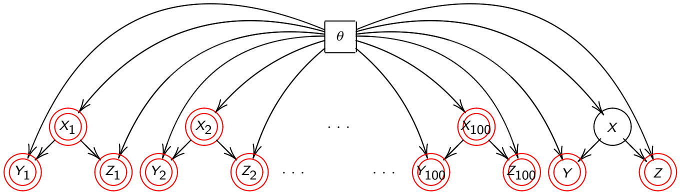


global models



prior ignorance
about $\theta \in [0, 1]^5$

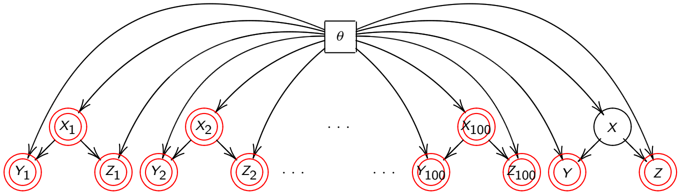
global models



prior ignorance
about $\theta \in [0,1]^5$

302 variables
observed

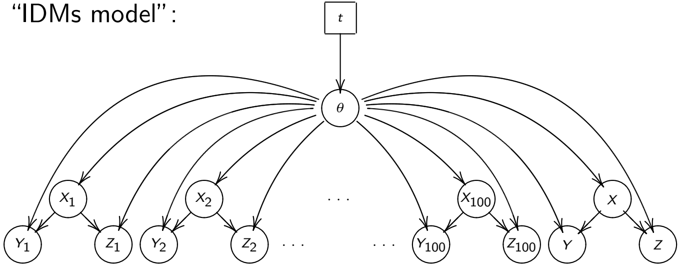
global models



prior ignorance
about $\theta \in [0, 1]^5$

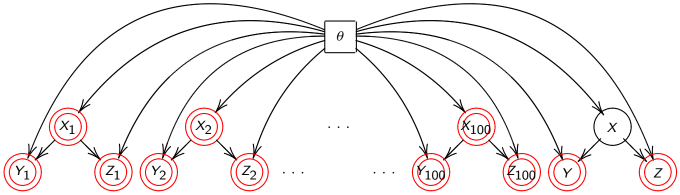
302 variables
observed

"IDMs model":



prior ignorance
about $t \in [0, 1]^5$

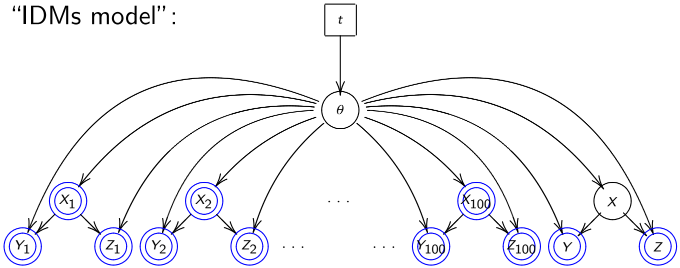
global models



prior ignorance about $\theta \in [0, 1]^5$

302 variables observed

“IDMs model”:



prior ignorance about $t \in [0, 1]^5$

302 variables observed