Efficient Algorithms for Bayesian Network Parameter Learning from Incomplete Data

Guy Van den Broeck, Karthika Mohan, Arthur Choi, Adnan Darwiche, and Judea Pearl

UCLA

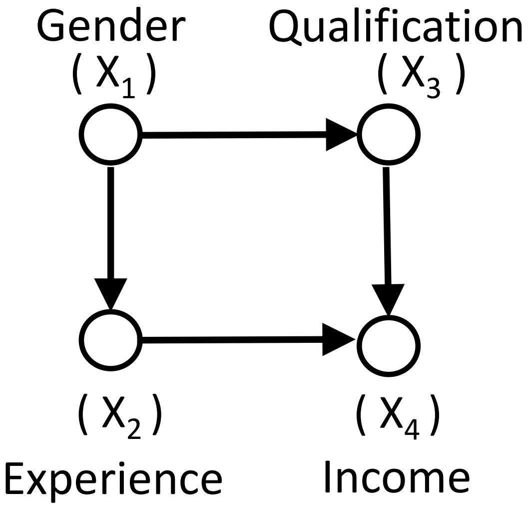
UAI 2015

Learning from Incomplete Data

- Input: data and BN structure

  E.g., Gender wage gap study

X1 (Gender)	X2 (Experience)	X3 (Qualification)	X4 (Income)
0	1	0	1
1	1	?	1
0	1	0	1
1	?	1	0
1	0	?	?
0	0	?	?
0	1	0	1



- Output: BN parameters

  E.g., θ_{Gender}, $\theta_{Experience|Gender}$, $\theta_{Qualification|Gender,}$, etc.

Current Approaches: Properties

	Likelihood Optimization
Inference-Free	✗
Consistent for MCAR	✔
Consistent for MAR	✔
Consistent for MNAR	✗
Maximum Likelihood	✔

Current Approaches: Properties

	Likelihood Optimization	Expectation Maximization
Inference-Free	✘	✘
Consistent for MCAR	✔	✔ / ✘
Consistent for MAR	✔	✔ / ✘
Consistent for MNAR	✘	✘
Maximum Likelihood	✔	✔ / ✘
Closed Form	n/a	✘
Passes over the data	n/a	?

Problem Statement

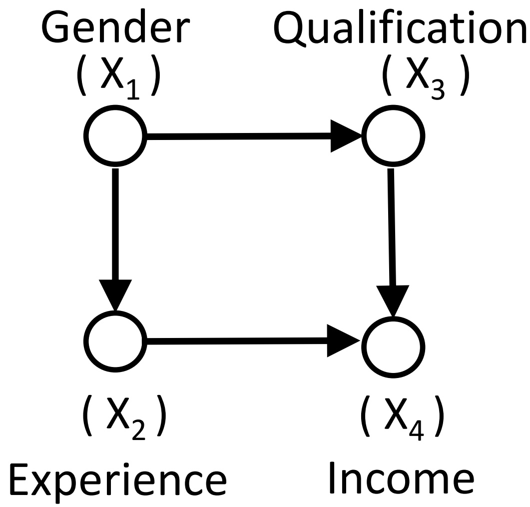
	Likelihood Optimization	Expectation Maximization
Inference-Free	✘	✘
Consistent for MCAR	✔	✔/✘
Consistent for MAR	✔	✔/✘
Consistent for MNAR	✘	✘
Maximum Likelihood	✔	✔/✘
Closed Form	n/a	✘
Passes over the data	n/a	?

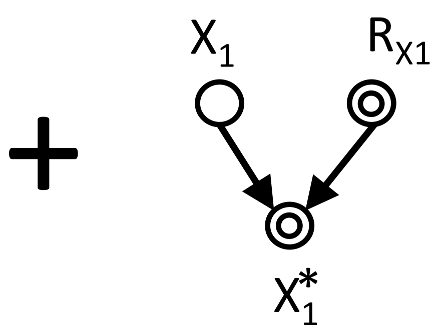
Conventional wisdom: this is inevitable!

Contribution

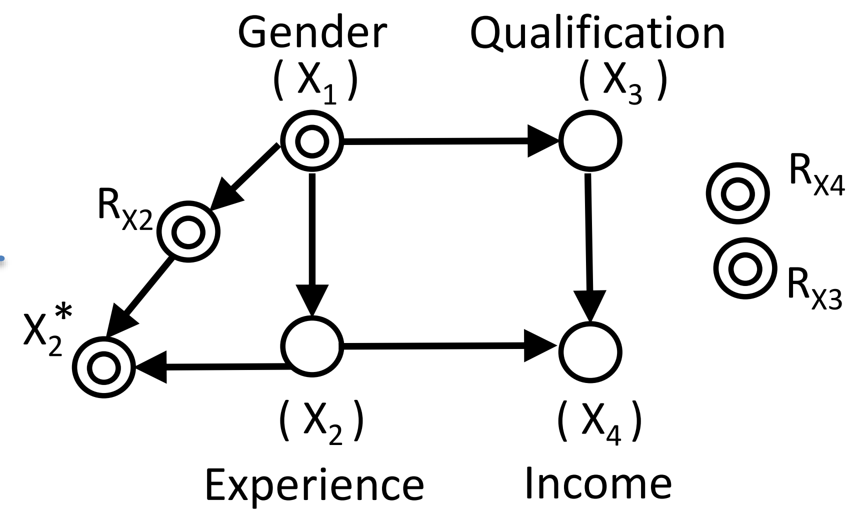
	Likelihood Optimization	Expectation Maximization	Deletion [this paper]
Inference-Free	✗	✗	✔
Consistent for MCAR	✔	✔/✗	✔
Consistent for MAR	✔	✔/✗	✔
Consistent for MNAR	✗	✗	✔/✗
Maximum Likelihood	✔	✔/✗	✗
Closed Form	n/a	✗	✔
Passes over the data	n/a	?	1

Missingness Graphs





$$X_1^* = \begin{cases} X_1 & \text{if } R_{X1} = \text{ob} \\ m & \text{if } R_{X1} = \text{unob} \end{cases}$$



Fully observed variables

$\mathbf{X}_o = \{X_1\}$

Partially observed variables

$\mathbf{X}_m = \{X_2, X_3, X_4\}$

Missingness Dataset

- Encoding of the data
  - Fully observed vars $\mathbf{X}_o$
  - Causal mechanisms $\mathbf{R}$
  - Proxies for $\mathbf{X}_m$

$$X_1^* = \begin{cases} X_1 & \text{if } R_{X1} = ob \\ \color{red}{m} & \text{if } R_{X1} = unob \end{cases}$$

- Fully observed
- Data distribution $Pr_D(.)$

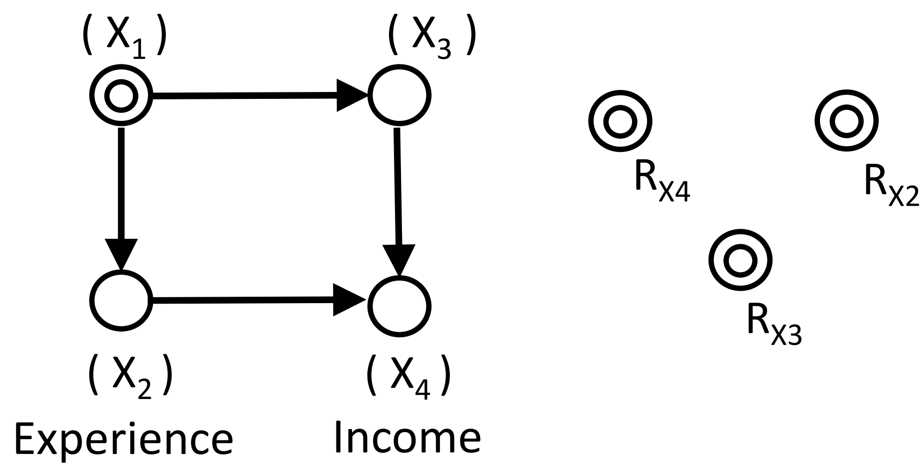
X_1	X^*_2	X^*_3	R_{X2}	R_{X3}	P^*
0	0	0	ob	ob	0.200
0	0	1	ob	ob	0.100
0	1	0	ob	ob	0.050
0	1	1	ob	ob	0.050
1	0	0	ob	ob	0.060
1	0	1	ob	ob	0.040
1	1	0	ob	ob	0.070
1	1	1	ob	ob	0.030
0	0	m	ob	unob	0.100
0	1	m	ob	unob	0.020
1	0	m	ob	unob	0.080
1	1	m	ob	unob	0.180
0	m	0	unob	ob	0.100
0	m	1	unob	ob	0.020
…	…	…	…	…	…

Algorithms

- Missingness categories (classes of graphs)
  - Missing Completely At Random (MCAR)
  - Missing At Random (MAR)
  - Missing Not At Random (MNAR)
- Deletion techniques
  - Direct Deletion
  - Factored Deletion
  - Informed Deletion

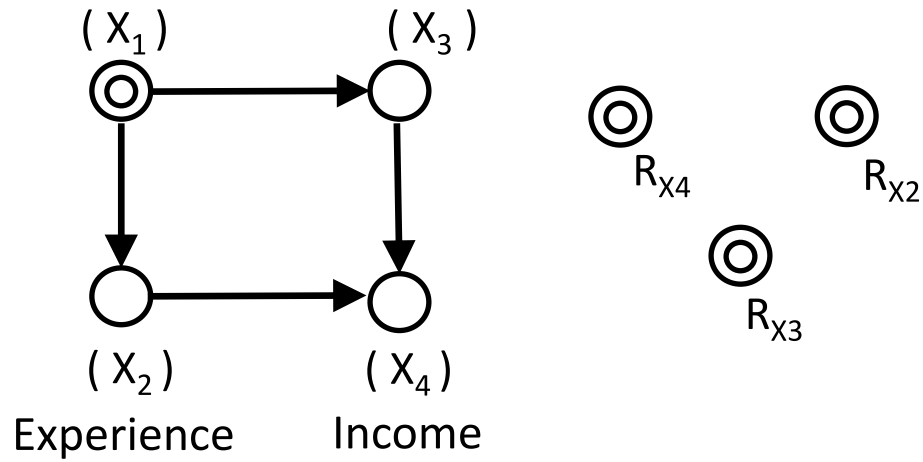
Missing Completely at Random (MCAR)

$$(\mathbf{X}_m\ \mathbf{X}_o) \perp\!\!\!\perp \mathbf{R}$$



Missing Completely at Random (MCAR)

$(\mathbf{X}_m \, \mathbf{X}_o) \perp\!\!\!\perp \mathbf{R}$



$$(X_1 \, X_2 \, X_3 \, X_4) \perp\!\!\!\perp (R_{X2} \, R_{X3} \, R_{X4})$$

Direct Deletion (MCAR)

$(\mathbf{X}_m\ \mathbf{X}_o) \perp\!\!\!\perp \mathbf{R}$

Independencies:

- $(X_1\ X_2) \perp\!\!\!\perp R$
- $(X_1\ X_2) \perp\!\!\!\perp R_{X2}$

Estimand:

$$Pr(X_1, X_2)$$

Direct Deletion (MCAR)

$(X_m\ X_o) \perp\!\!\!\perp R$

Independencies:

- $(X_1\ X_2) \perp\!\!\!\perp R$
- $(X_1\ X_2) \perp\!\!\!\perp R_{X2}$

Estimand:

$$Pr(X_1, X_2)$$
$$= Pr\left(X_1 X_2 \middle| R_{X_2} = ob\right)$$

Direct Deletion (MCAR)

$(X_m\ X_o) \perp\!\!\!\perp R$

Independencies:

- $(X_1\ X_2) \perp\!\!\!\perp R$
- $(X_1\ X_2) \perp\!\!\!\perp R_{X2}$

Estimand:

$$
\begin{aligned}
&Pr(X_1, X_2) \\
&\quad = Pr\left(X_1 X_2 \middle| R_{X_2} = ob\right) \\
&\quad = Pr\left(X_1 X_2^* \middle| R_{X_2} = ob\right)
\end{aligned}
$$

Direct Deletion (MCAR)

(X_m X_o) ⫫ R

Independencies:

- $(X_1 X_2) \perp\!\!\!\perp R$
- $(X_1 X_2) \perp\!\!\!\perp R_{X2}$

Estimand:

$$
\begin{aligned}
Pr(X_1, X_2) \\
&= Pr\left(X_1 X_2 \middle| R_{X_2} = ob\right) \\
&= Pr\left(X_1 X_2^* \middle| R_{X_2} = ob\right) \\
&= Pr_D(X_1 X_2^* | R_{X_2} = ob)
\end{aligned}
$$

Direct Deletion (MCAR)

$(X_m X_o) \perp\!\!\!\perp R$

Independencies:

- $(X_1 X_2) \perp\!\!\!\perp R$
- $(X_1 X_2) \perp\!\!\!\perp R_{X2}$

Estimand:

$$Pr(X_1, X_2)$$
$$= Pr\left(X_1 X_2 \middle| R_{X_2} = ob\right)$$
$$= Pr\left(X_1 X_2^* \middle| R_{X_2} = ob\right)$$
$$= Pr_D(X_1 X_2^* | R_{X_2} = ob)$$

X_1	X^*_2	X^*_3	R_{X2}	R_{X3}	P*
0	0	0	ob	ob	0.200
0	0	1	ob	ob	0.100
0	1	0	ob	ob	0.050
0	1	1	ob	ob	0.050
…	…	…	…	…	…
0	1	m	ob	unob	0.020
1	0	m	ob	unob	0.080
1	1	m	ob	unob	0.180
~~0~~	~~m~~	~~0~~	~~unob~~	~~ob~~	~~0.100~~
~~0~~	~~m~~	~~1~~	~~unob~~	~~ob~~	~~0.020~~
…	…	…	…	…	…

Direct Deletion (MCAR)

$(X_m\ X_o) \perp\!\!\!\perp R$

Independencies:

- $(X_1\ X_2) \perp\!\!\!\perp R$
- $(X_1\ X_2) \perp\!\!\!\perp R_{X2}$

Estimand:

$$Pr(X_1, X_2)$$
$$= Pr\left(X_1 X_2 \middle| R_{X_2} = ob\right)$$
$$= Pr\left(X_1 X_2^* \middle| R_{X_2} = ob\right)$$
$$= Pr_D(X_1 X_2^* | R_{X_2} = ob)$$

X_1	X^*_2	X^*_3	R_{X2}	R_{X3}	P*
0	0	0	ob	ob	0.200
0	0	1	ob	ob	0.100
0	1	0	ob	ob	0.050
0	1	1	ob	ob	0.050
…	…	…	…	…	…
0	1	m	ob	unob	0.020
1	0	m	ob	unob	0.080
1	1	m	ob	unob	0.180
~~0~~	~~m~~	~~0~~	~~unob~~	~~ob~~	~~0.100~~
~~0~~	~~m~~	~~1~~	~~unob~~	~~ob~~	~~0.020~~
…	…	…	…	…	…

Cf. *listwise* and *pairwise* deletion in statistics

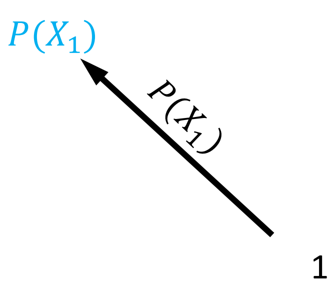
Factored Deletion (MCAR)

Many ways of factorizing the estimand!

X_1	X^*_2	X^*_3	R_{X2}	R_{X3}	P*
0	0	0	ob	ob	0.200
0	0	1	ob	ob	0.100
0	1	0	ob	ob	0.050
0	1	1	ob	ob	0.050
1	0	0	ob	ob	0.060
1	0	1	ob	ob	0.040
1	1	0	ob	ob	0.070
1	1	1	ob	ob	0.030
0	0	m	ob	unob	0.100
0	1	m	ob	unob	0.020
1	0	m	ob	unob	0.080
1	1	m	ob	unob	0.180
1	m	m	unob	unob	0.020

Factored Deletion (MCAR)

Many ways of factorizing the estimand!

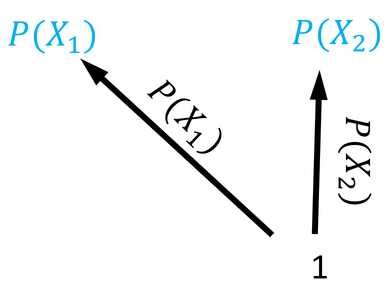


X_1	X^*_2	X^*_3	R_{X2}	R_{X3}	P*
0	0	0	ob	ob	0.200
0	0	1	ob	ob	0.100
0	1	0	ob	ob	0.050
0	1	1	ob	ob	0.050
1	0	0	ob	ob	0.060
1	0	1	ob	ob	0.040
1	1	0	ob	ob	0.070
1	1	1	ob	ob	0.030
0	0	m	ob	unob	0.100
0	1	m	ob	unob	0.020
1	0	m	ob	unob	0.080
1	1	m	ob	unob	0.180
1	m	m	unob	unob	0.020

Factored Deletion (MCAR)

Many ways of factorizing the estimand!

$P(X_2) = P(X_2|R_{X_2} = ob)$

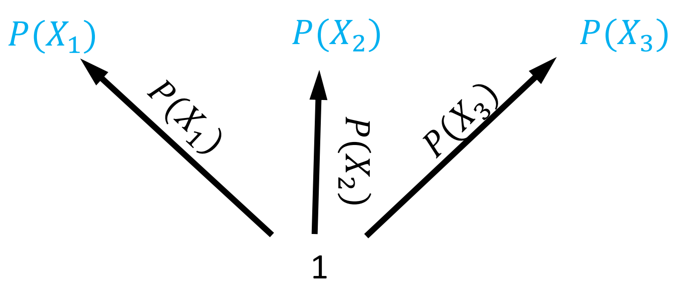


X_1	X^*_2	X^*_3	R_{X2}	R_{X3}	P*
0	0	0	ob	ob	0.200
0	0	1	ob	ob	0.100
0	1	0	ob	ob	0.050
0	1	1	ob	ob	0.050
1	0	0	ob	ob	0.060
1	0	1	ob	ob	0.040
1	1	0	ob	ob	0.070
1	1	1	ob	ob	0.030
0	0	m	ob	unob	0.100
0	1	m	ob	unob	0.020
1	0	m	ob	unob	0.080
1	1	m	ob	unob	0.180
1	m	m	unob	unob	0.020

Factored Deletion (MCAR)

Many ways of factorizing the estimand!

$P(X_3) = P(X_3|R_{X_3} = ob)$

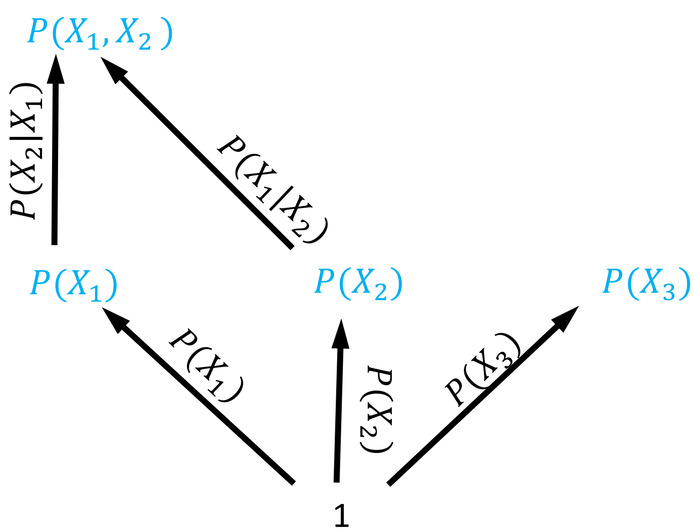


X_1	X^*_2	X^*_3	R_{X2}	R_{X3}	P*
0	0	0	ob	ob	0.200
0	0	1	ob	ob	0.100
0	1	0	ob	ob	0.050
0	1	1	ob	ob	0.050
1	0	0	ob	ob	0.060
1	0	1	ob	ob	0.040
1	1	0	ob	ob	0.070
1	1	1	ob	ob	0.030
0	0	m	ob	unob	0.100
0	1	m	ob	unob	0.020
1	0	m	ob	unob	0.080
1	1	m	ob	unob	0.180
1	m	m	unob	unob	0.020

Factored Deletion (MCAR)

Many ways of factorizing the estimand!

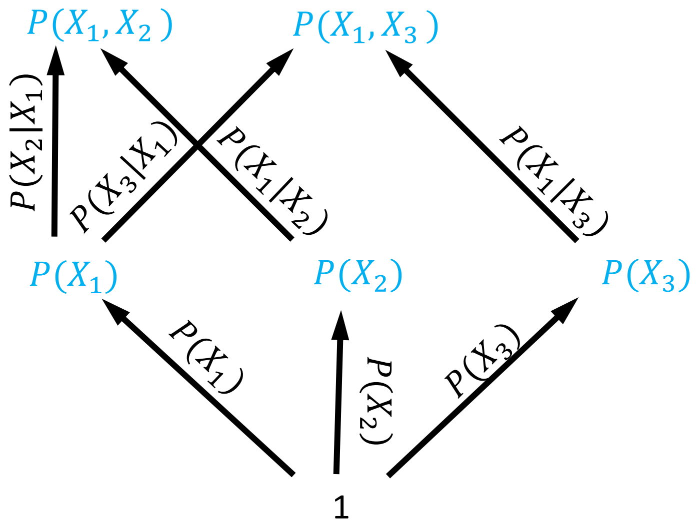
$$P(X_1, X_2) = P\big(X_2 | X_1, R_{X_2} = ob\big) P(X_1)$$
$$P(X_1, X_2) = P(X_1 | X_2, R_{X_2} = ob)\ P(X_2 | R_{X_2} = ob)$$



X_1	X^*_2	X^*_3	R_{X2}	R_{X3}	P*
0	0	0	ob	ob	0.200
0	0	1	ob	ob	0.100
0	1	0	ob	ob	0.050
0	1	1	ob	ob	0.050
1	0	0	ob	ob	0.060
1	0	1	ob	ob	0.040
1	1	0	ob	ob	0.070
1	1	1	ob	ob	0.030
0	0	m	ob	unob	0.100
0	1	m	ob	unob	0.020
1	0	m	ob	unob	0.080
1	1	m	ob	unob	0.180
1	m	m	unob	unob	0.020

Factored Deletion (MCAR)

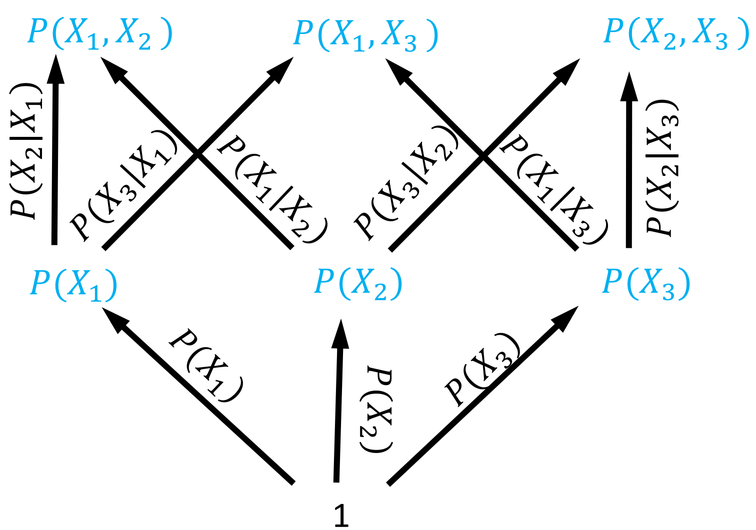
Many ways of factorizing the estimand!



X_1	X^*_2	X^*_3	R_{X2}	R_{X3}	P*
0	0	0	ob	ob	0.200
0	0	1	ob	ob	0.100
0	1	0	ob	ob	0.050
0	1	1	ob	ob	0.050
1	0	0	ob	ob	0.060
1	0	1	ob	ob	0.040
1	1	0	ob	ob	0.070
1	1	1	ob	ob	0.030
0	0	m	ob	unob	0.100
0	1	m	ob	unob	0.020
1	0	m	ob	unob	0.080
1	1	m	ob	unob	0.180
1	m	m	unob	unob	0.020

Factored Deletion (MCAR)

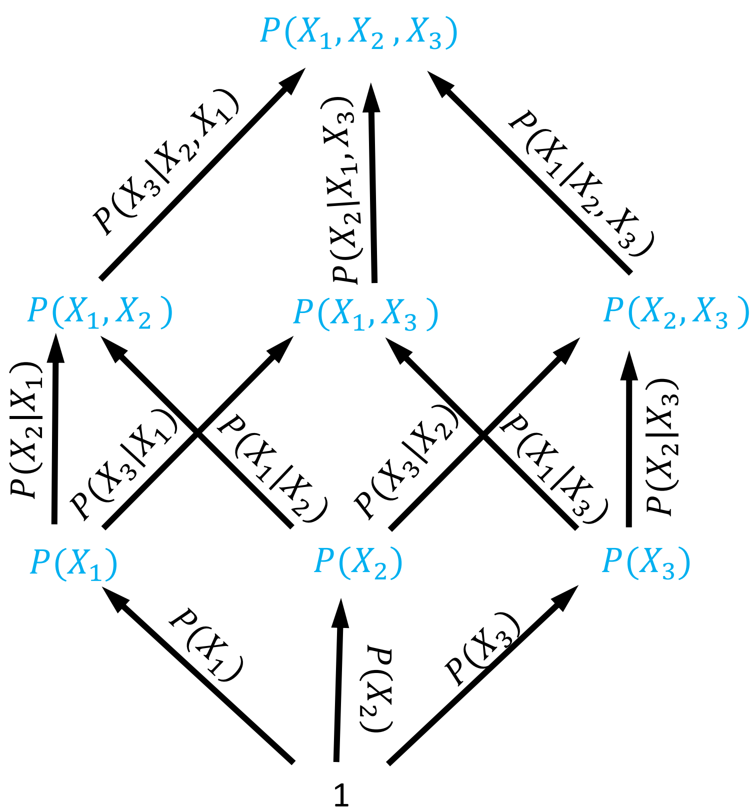
Many ways of factorizing the estimand!



X_1	X^*_2	X^*_3	R_{X2}	R_{X3}	P*
0	0	0	ob	ob	0.200
0	0	1	ob	ob	0.100
0	1	0	ob	ob	0.050
0	1	1	ob	ob	0.050
1	0	0	ob	ob	0.060
1	0	1	ob	ob	0.040
1	1	0	ob	ob	0.070
1	1	1	ob	ob	0.030
0	0	m	ob	unob	0.100
0	1	m	ob	unob	0.020
1	0	m	ob	unob	0.080
1	1	m	ob	unob	0.180
1	m	m	unob	unob	0.020

Factored Deletion (MCAR)

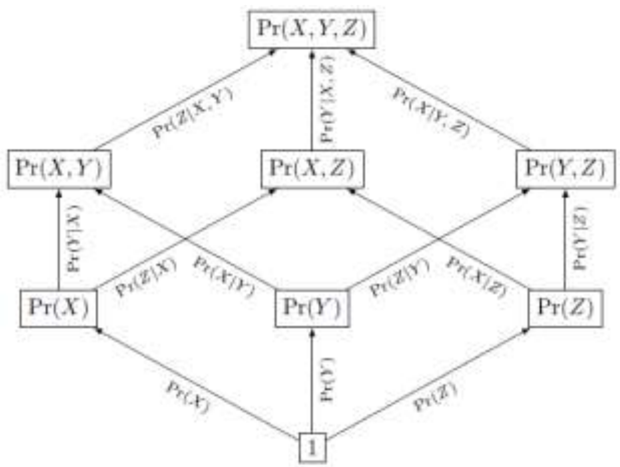
Many ways of factorizing the estimand!



X_1	X^*_2	X^*_3	R_{X2}	R_{X3}	P*
0	0	0	ob	ob	0.200
0	0	1	ob	ob	0.100
0	1	0	ob	ob	0.050
0	1	1	ob	ob	0.050
1	0	0	ob	ob	0.060
1	0	1	ob	ob	0.040
1	1	0	ob	ob	0.070
1	1	1	ob	ob	0.030
0	0	m	ob	unob	0.100
0	1	m	ob	unob	0.020
1	0	m	ob	unob	0.080
1	1	m	ob	unob	0.180
1	m	m	unob	unob	0.020

Factored Deletion (MCAR)

- Aggregate all factorizations in lattice
- Simple algorithm

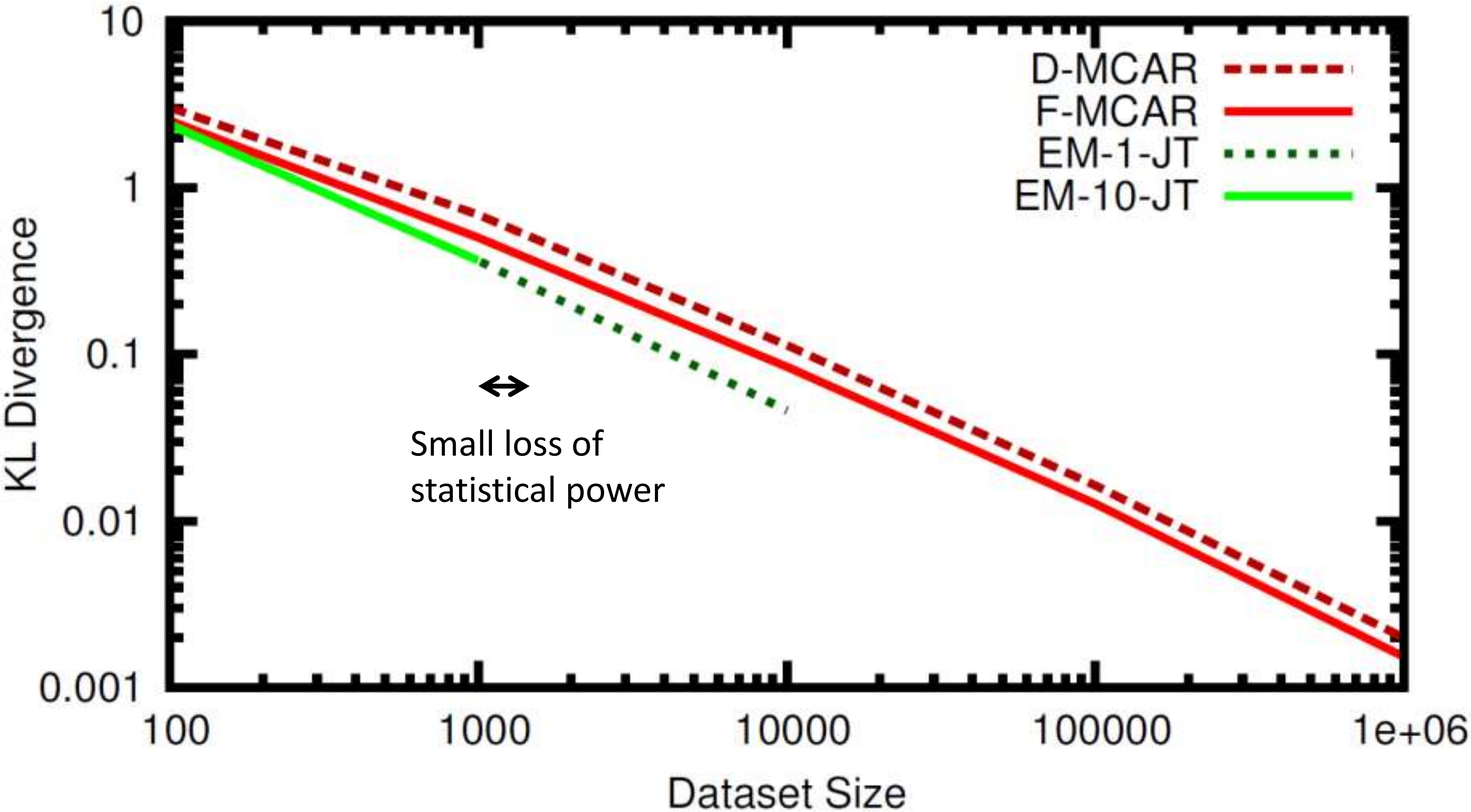


Algorithm 1 F-MCAR($\mathbf{y}, \mathcal{D}$)

Input:
$\mathbf{y}$: A state of query variables $\mathbf{Y}$
$\mathcal{D}$: An incomplete dataset with data distribution $\Pr_{\mathcal{D}}$
Auxiliary:
CACHE: A global cache of estimated probabilities
Function:
if $\mathbf{y} = \emptyset$ **then return** 1
if CACHE[$\mathbf{y}$] $\neq nil$ **then return** CACHE[$\mathbf{y}$]
$\mathcal{E} \leftarrow \emptyset$ // Initialize set of estimates
for each $y \in \mathbf{y}$ **do**
  $\mathbf{u} \leftarrow \mathbf{y} \setminus \{y\}$ // Factorize with parents $\mathbf{u}$
  add $\Pr_{\mathcal{D}}(y|\mathbf{u}, \mathbf{R}_{\mathbf{y}}{=}\text{ob}) \cdot$ F-MCAR($\mathbf{u}, \mathcal{D}$) **to** $\mathcal{E}$
CACHE[$\mathbf{y}$] ←Aggregate estimates in $\mathcal{E}$ // E.g., mean
return CACHE[$\mathbf{y}$]

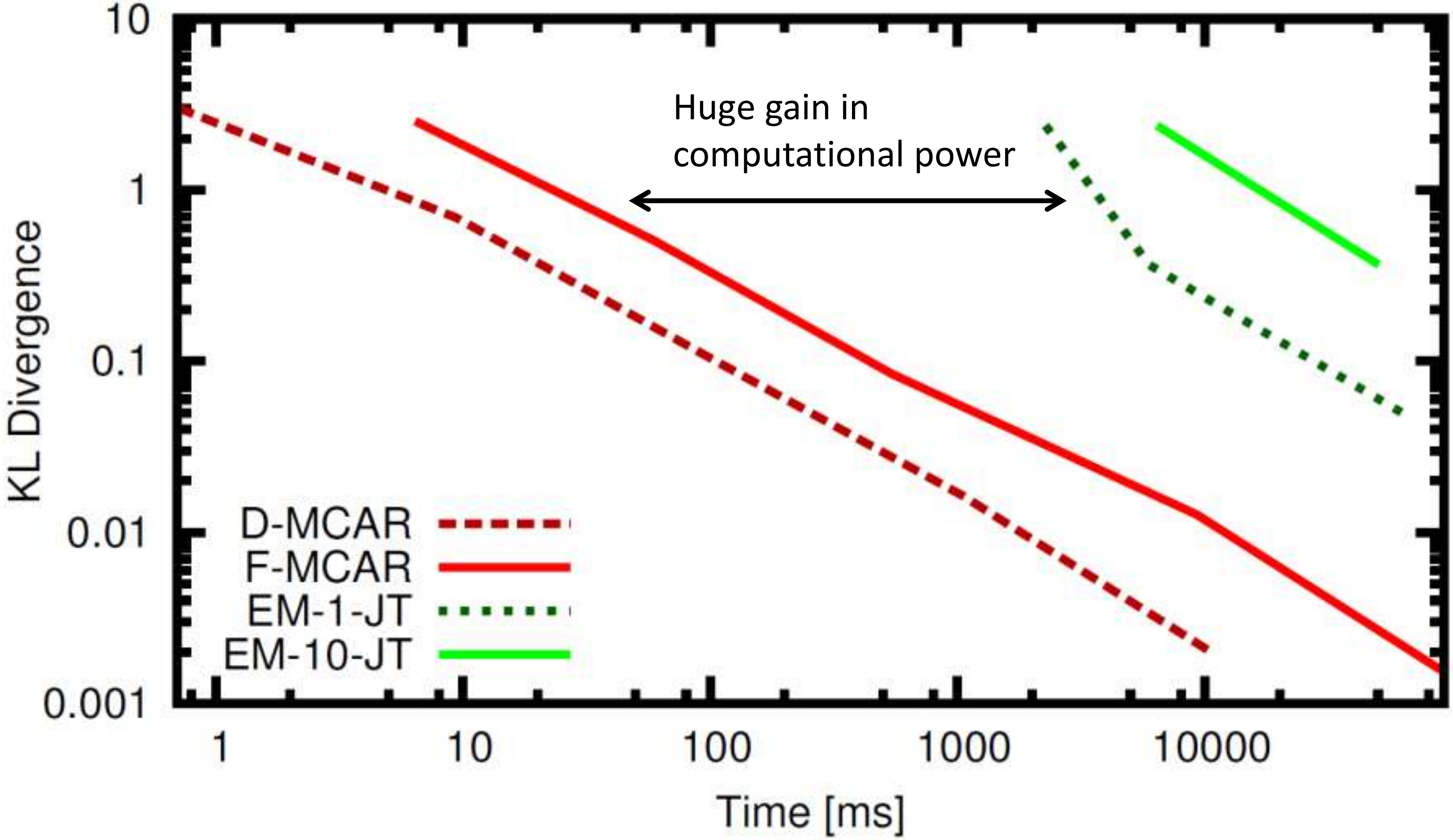
- Data used for Pr(Y)?
  - Direct deletion: data where **all** Y observed
  - Factored deletion: data where **some** Y observed

MCAR Experiments (data size)



(Alarm network)

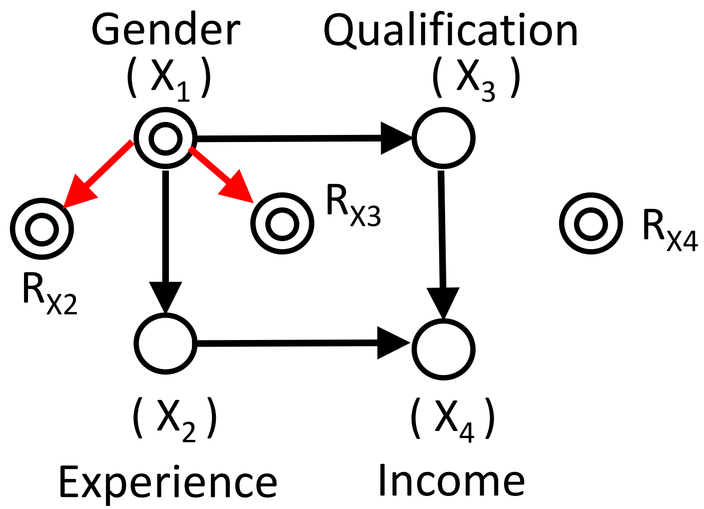
MCAR Experiments (time)



(Alarm network)

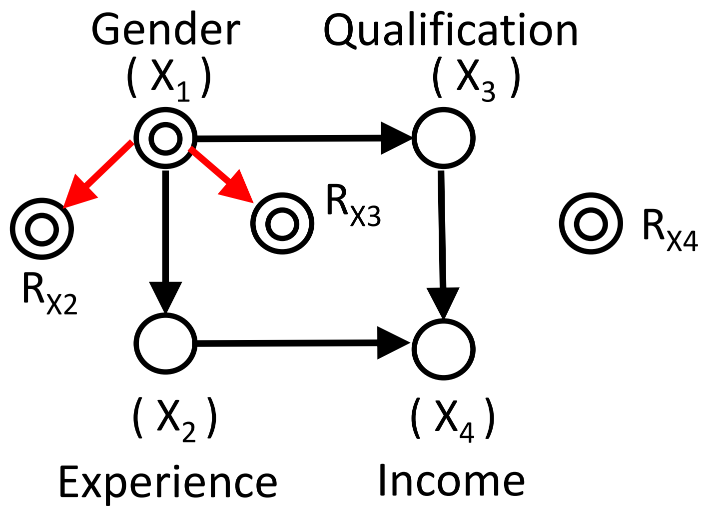
Missing At Random (MAR)

$\mathbf{X}_m \perp\!\!\!\perp \mathbf{R} \mid \mathbf{X}_o$



Missing At Random (MAR)

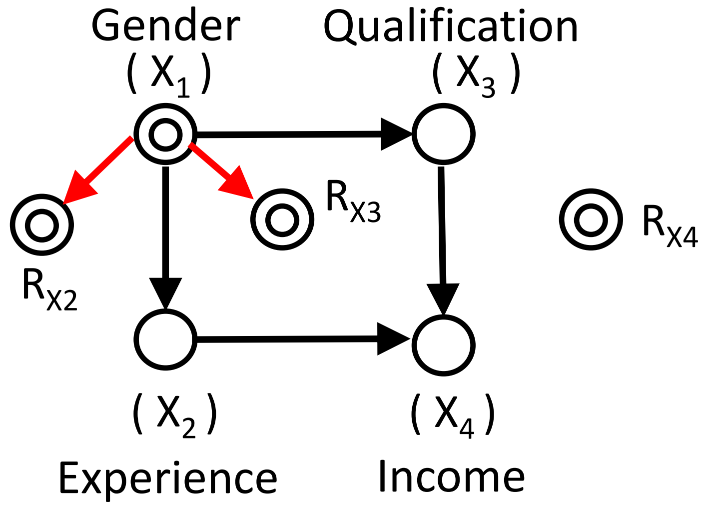
$$\mathbf{X}_m \perp\!\!\!\perp \mathbf{R} \mid \mathbf{X}_o$$



$$(X_2\ X_3\ X_4) \perp\!\!\!\perp (R_{X2}\ R_{X3}\ R_{X4}) \mid X_1$$

Missing At Random (MAR)

$$\mathbf{X}_m \perp\!\!\!\perp \mathbf{R} \mid \mathbf{X}_o$$



$$(X_2\ X_3\ X_4) \perp\!\!\!\perp (R_{X2}\ R_{X3}\ R_{X4}) \mid X_1$$

Most-general class where maximum-likelihood is consistent!

Direct Deletion (MAR)

$X_m \perp\!\!\!\perp R \mid X_o$

Independencies:

$(X_2\ X_3) \perp\!\!\!\perp R \mid X_1$

Estimand:

X_1	X^*_2	X^*_3	R_{X2}	R_{X3}	P*
0	0	0	ob	ob	0.200
0	0	1	ob	ob	0.100
0	1	0	ob	ob	0.050
0	1	1	ob	ob	0.050
1	0	0	ob	ob	0.060
1	0	1	ob	ob	0.040
1	1	0	ob	ob	0.070
1	1	1	ob	ob	0.030
0	0	m	ob	unob	0.100
0	1	m	ob	unob	0.020
1	0	m	ob	unob	0.080
1	1	m	ob	unob	0.180
1	m	m	unob	unob	0.020

Direct Deletion (MAR)

$X_m \perp\!\!\!\perp R \mid X_o$

Independencies:

$(X_2\ X_3) \perp\!\!\!\perp R \mid X_1$

Estimand:

$$P(X_2, X_3)$$
$$= \sum_{X_1} P(X_2 X_3 | X_1) P(X_1)$$
$$= \sum_{X_1} P(X_2 X_3 | X_1, R = ob) P(X_1)$$

X_1	X^*_2	X^*_3	R_{X2}	R_{X3}	P*
0	0	0	ob	ob	0.200
0	0	1	ob	ob	0.100
0	1	0	ob	ob	0.050
0	1	1	ob	ob	0.050
1	0	0	ob	ob	0.060
1	0	1	ob	ob	0.040
1	1	0	ob	ob	0.070
1	1	1	ob	ob	0.030
0	0	m	ob	unob	0.100
0	1	m	ob	unob	0.020
1	0	m	ob	unob	0.080
1	1	m	ob	unob	0.180
1	m	m	unob	unob	0.020

Direct Deletion (MAR)

$X_m \perp\!\!\!\perp R \mid X_o$

Independencies:

$(X_2\ X_3) \perp\!\!\!\perp R \mid X_1$

Estimand:

$$P(X_2, X_3)$$
$$= \sum_{X_1} P(X_2X_3|X_1)P(X_1)$$
$$= \sum_{X_1} \boxed{P(X_2X_3|X_1, R = ob)} P(X_1)$$

X_1	X^*_2	X^*_3	R_{X2}	R_{X3}	P*
0	0	0	ob	ob	0.200
0	0	1	ob	ob	0.100
0	1	0	ob	ob	0.050
0	1	1	ob	ob	0.050
1	0	0	ob	ob	0.060
1	0	1	ob	ob	0.040
1	1	0	ob	ob	0.070
1	1	1	ob	ob	0.030
0	0	m	ob	unob	0.100
0	1	m	ob	unob	0.020
1	0	m	ob	unob	0.080
1	1	m	ob	unob	0.180
1	m	m	unob	unob	0.020

Direct Deletion (MAR)

$X_m \perp\!\!\!\perp R \mid X_o$

Independencies:

$(X_2\ X_3) \perp\!\!\!\perp R \mid X_1$

Estimand:

$$P(X_2, X_3)$$
$$= \sum_{X_1} P(X_2 X_3 | X_1) P(X_1)$$
$$= \sum_{X_1} \boxed{P(X_2 X_3 | X_1, R = ob)} P(X_1)$$

X_1	X^*_2	X^*_3	R_{X2}	R_{X3}	P*
0	0	0	ob	ob	0.200
0	0	1	ob	ob	0.100
0	1	0	ob	ob	0.050
0	1	1	ob	ob	0.050
1	0	0	ob	ob	0.060
1	0	1	ob	ob	0.040
1	1	0	ob	ob	0.070
1	1	1	ob	ob	0.030
0	0	m	ob	unob	0.100
0	1	m	ob	unob	0.020
1	0	m	ob	unob	0.080
1	1	m	ob	unob	0.180
1	m	m	unob	unob	0.020

Factored Deletion (MAR)

Algorithm 2 F-MAR($\mathbf{y}, \mathcal{D}$)

Input:

$\mathbf{y}$: A state of query variables $\mathbf{Y}$, consisting of $\mathbf{y}_o$ and $\mathbf{y}_m$

$\mathcal{D}$: An incomplete dataset with data distribution $\Pr_{\mathcal{D}}$

Function:

```
e ← 0                                          // Estimated probability
for each x_o appearing in D that agrees with y_o do
    D_{x_o} ← subset of D where x_o holds
    e ← e + Pr_D(x_o) · F-MCAR(y_m, D_{x_o})
return e
```

$e \leftarrow 0$ // Estimated probability

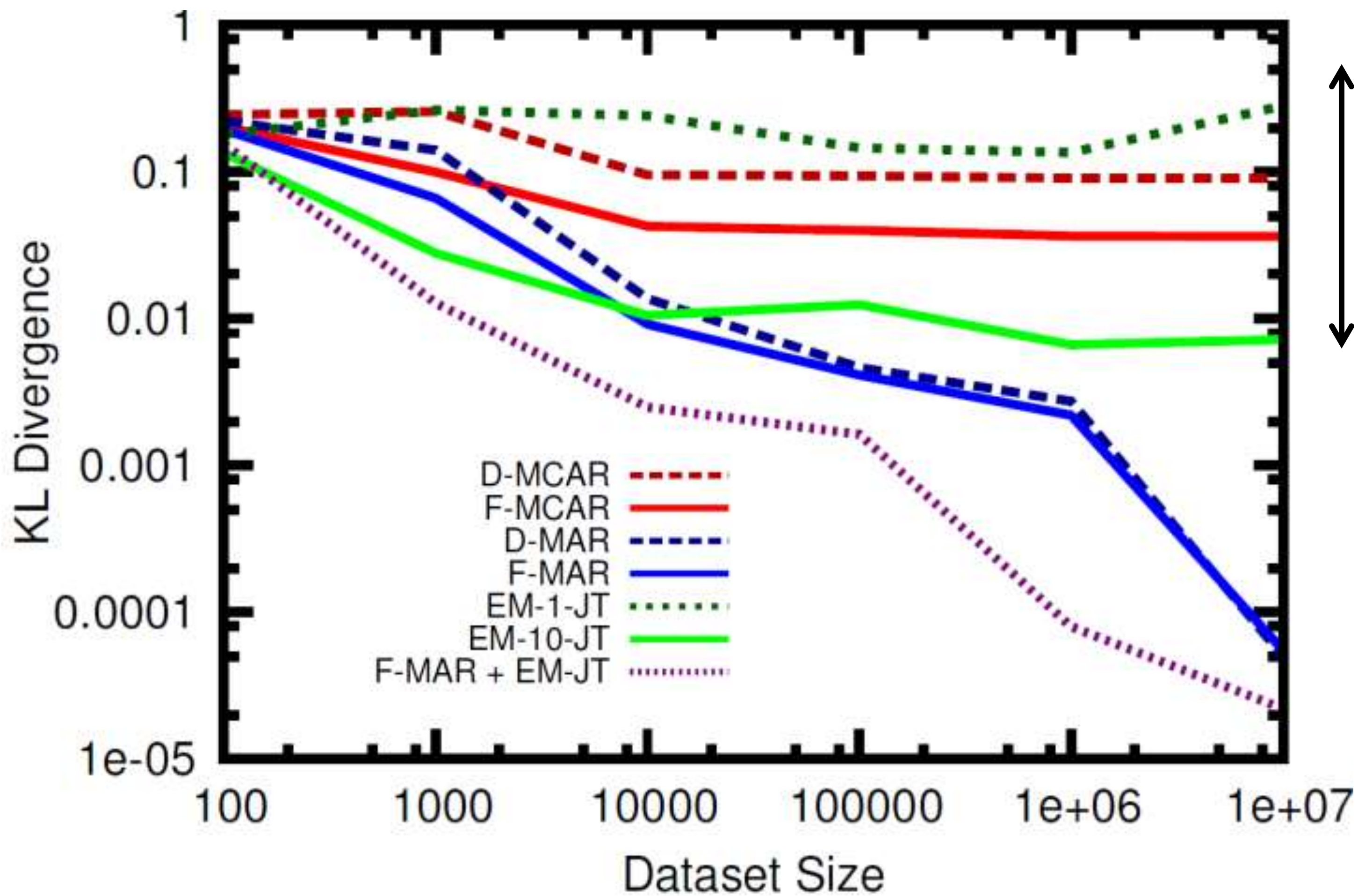
for each $\mathbf{x}_o$ appearing in $\mathcal{D}$ that agrees with $\mathbf{y}_o$ **do**

$\quad \mathcal{D}_{\mathbf{x}_o} \leftarrow$ subset of $\mathcal{D}$ where $\mathbf{x}_o$ holds

$\quad e \leftarrow e + \Pr_{\mathcal{D}}(\mathbf{x}_o) \cdot \text{F-MCAR}(\mathbf{y}_m, \mathcal{D}_{\mathbf{x}_o})$

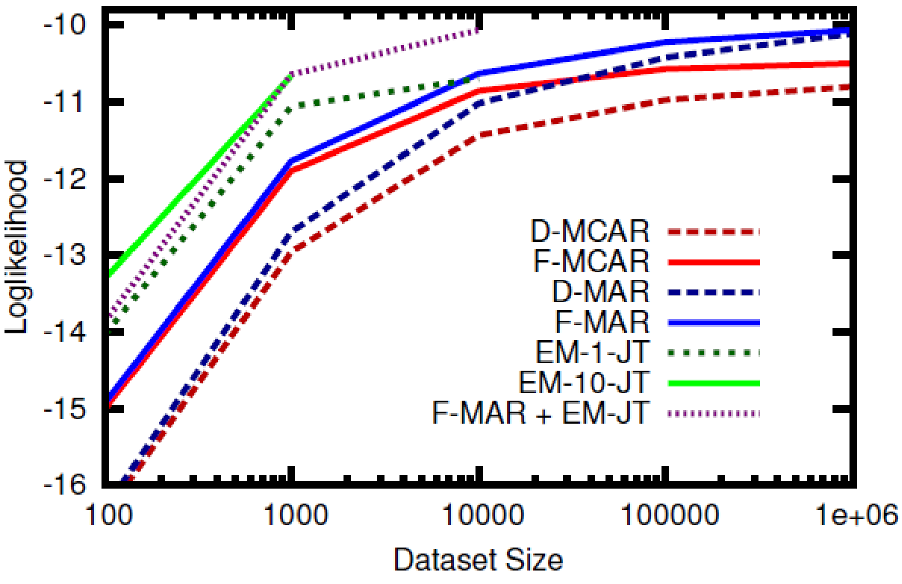
return e

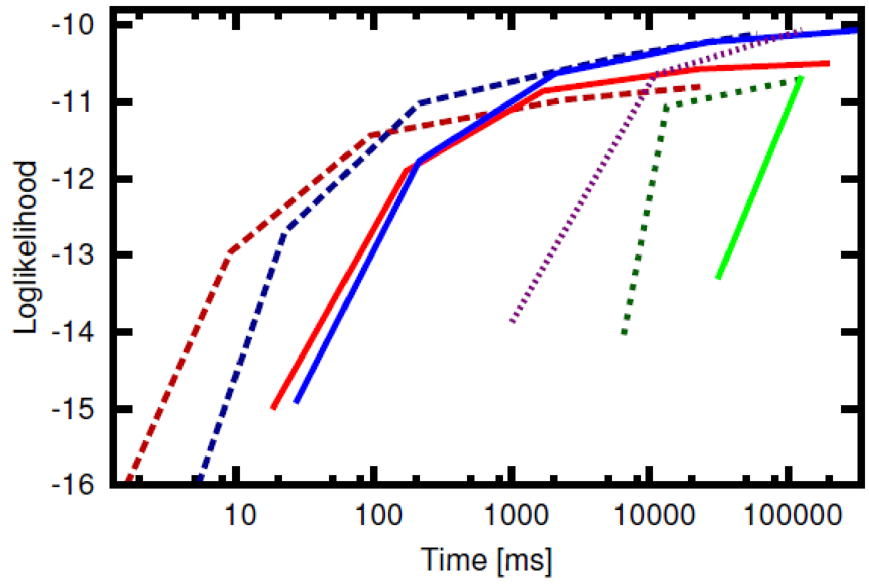
MAR Experiments (Fire Alarm)



INCONSISTENT

MAR Experiments (Alarm)





MAR Experiments (Intractable)

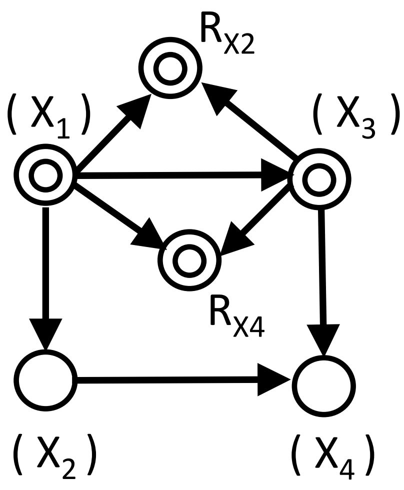
Log-likelihoods of large intractable networks

Size		EM-JT	EM-BP	D-MCAR	F-MCAR	D-MAR	F-MAR		EM-JT	EM-BP	D-MCAR	F-MCAR	D-MAR	F-MAR
10^2	Grid 90-20-1	-	-57.14	-80.92	-57.01	-80.80	**-56.53**	Water	-19.10	**-18.76**	-25.31	-21.76	-25.29	-21.81
10^3	Grid 90-20-1	-	-65.41	-38.54	-30.07	-38.27	**-29.86**	Water	-	**-14.73**	-19.13	-16.45	-18.93	-16.36
10^4	Grid 90-20-1	-	-	-25.95	-23.30	-25.36	**-22.88**	Water	-	-20.70	-16.66	-14.90	-16.33	**-14.67**
10^5	Grid 90-20-1	-	-	-22.74	-22.01	**-21.60**	-	Water	-	-	-15.49	-	**-14.90**	-
10^2	Munin 1	-	**-103.72**	-115.50	-105.81	-115.41	-104.87	Barley	-	-89.22	-89.54	-89.26	-89.60	**-89.14**
10^3	Munin 1	-	-69.03	-71.01	-65.91	-70.61	**-65.51**	Barley	-	-74.26	-71.67	-70.46	-71.68	**-70.18**
10^4	Munin 1	-	-157.23	-56.07	**-54.24**	-55.46	-	Barley	-	-	-56.44	**-55.12**	-56.40	-
10^5	Munin 1	-	-	**-52.00**	-	-	-	Barley	-	-	-	-	-	-

Size		EM-JT	EM-BP	D-MCAR	F-MCAR	D-MAR	F-MAR		EM-JT	EM-BP	D-MCAR	F-MCAR	D-MAR	F-MAR
10^2	Grid 90-20-1	-	**-49.15**	-80.00	-56.45	-79.81	-55.94	Water	-18.88	**-18.73**	-25.84	-22.11	-25.87	-22.25
10^3	Grid 90-20-1	-	-53.64	-38.14	-29.32	-37.75	**-29.09**	Water	-17.63	**-14.41**	-18.39	-15.95	-18.27	-15.79
10^4	Grid 90-20-1	-	-85.65	-26.21	-23.05	-25.45	**-22.62**	Water	-	-14.52	-15.57	-14.07	-15.24	**-13.92**
10^5	Grid 90-20-1	-	-	-22.78	-21.54	-21.60	**-20.79**	Water	-	-24.99	-14.17	-13.46	-13.71	**-13.19**
10^6	Grid 90-20-1	-	-	-	-	-	-	Water	-	-	**-13.73**	-	-	-
10^2	Munin 1	-	**-99.15**	-114.76	-106.07	-114.66	-105.12	Barley	-89.05	-89.15	-89.57	-89.17	-89.62	**-89.03**
10^3	Munin 1	-	-67.85	-74.18	-67.81	-73.82	**-67.39**	Barley	-	-70.38	-71.86	-70.54	-71.87	**-70.27**
10^4	Munin 1	-	-66.62	-57.50	-54.94	-56.96	**-54.64**	Barley	-	-76.48	-56.37	**-55.13**	-56.33	-
10^5	Munin 1	-	-	-53.07	**-51.66**	-52.27	-	Barley	-	-	-51.31	-	**-51.19**	-

Informed Deletion

$\mathbf{X}_m \perp\!\!\!\perp \mathbf{R} \mid \mathbf{X}_o$



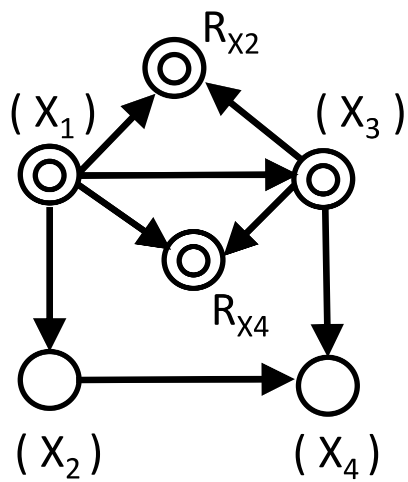
General m-graph depicting MAR

Direct Deletion

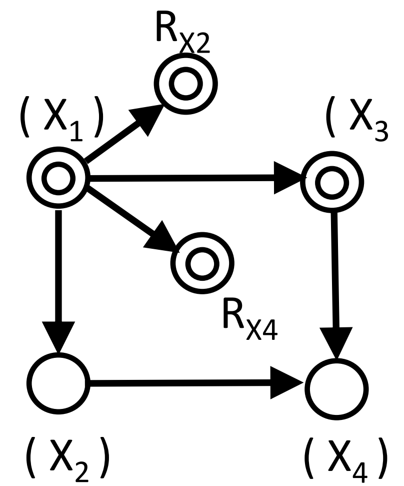
$$P(X_1, X_2) = \sum_{X_3} P\big(X_2 | X_1, X_3, R_{X_2} = ob\big) P(X_1, X_3)$$

Informed Deletion

$$\mathbf{X}_m \perp\!\!\!\perp \mathbf{R} \mid \mathbf{X}_o$$



General m-graph depicting MAR



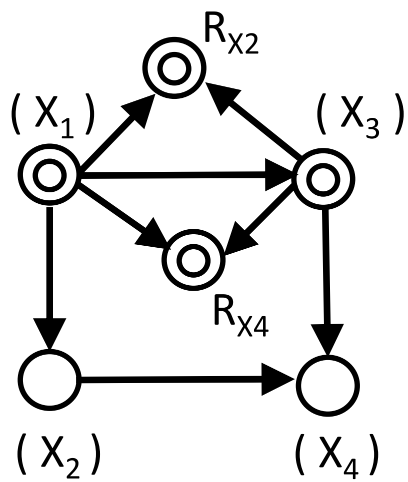
Problem specific m-graph

Direct Deletion

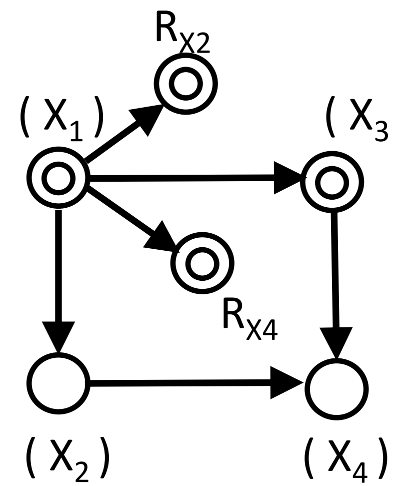
$$P(X_1, X_2) = \sum_{X_3} P\big(X_2 | X_1, X_3, R_{X_2} = ob\big) P(X_1, X_3)$$

Informed Deletion

$$\mathbf{X}_m \perp\!\!\!\perp \mathbf{R} \mid \mathbf{X}_o$$



General m-graph depicting MAR



Problem specific m-graph

$R_{X_2} \,||\, X_3 \mid X_1$

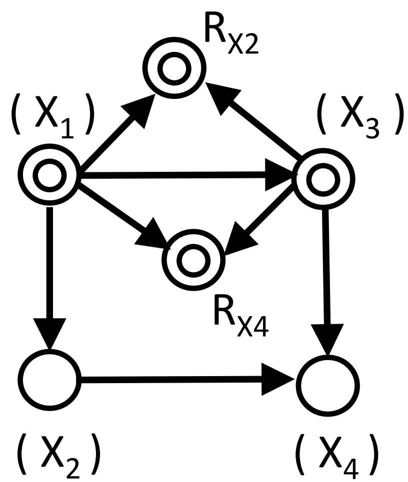
$R_{X_2} \,||\, X_1 \mid X_3$

Direct Deletion

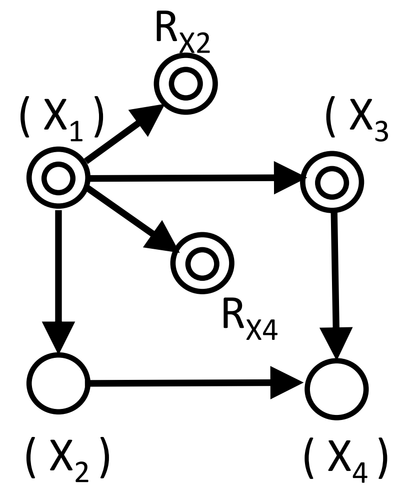
$$P(X_1, X_2) = \sum_{X_3} P\big(X_2 | X_1, X_3, R_{X_2} = ob\big) P(X_1, X_3)$$

Informed Deletion

$$\mathbf{X_m \perp\!\!\!\perp R \mid X_o}$$



General m-graph depicting MAR



Problem specific m-graph

$$\boldsymbol{R_{X_2} \,||\, X_3 \mid X_1}$$
$$\boldsymbol{R_{X_2} \,||\, X_1 \mid X_3}$$

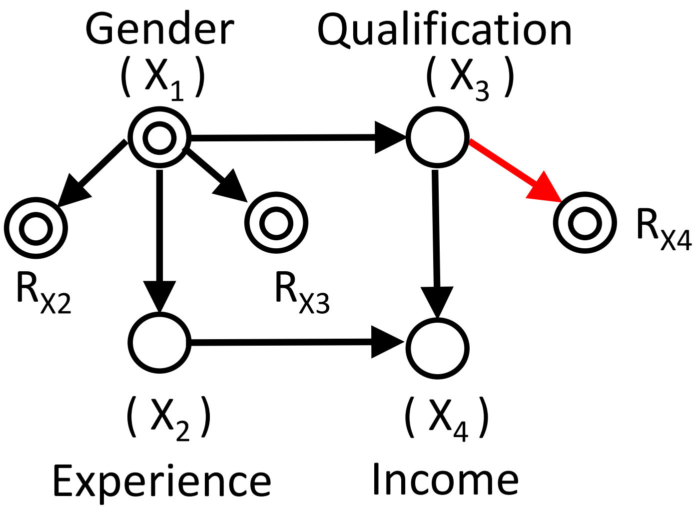
Direct Deletion

$$P(X_1, X_2) = \sum_{X_3} P\left(X_2 | X_1, X_3, R_{X_2} = ob\right) P(X_1, X_3)$$

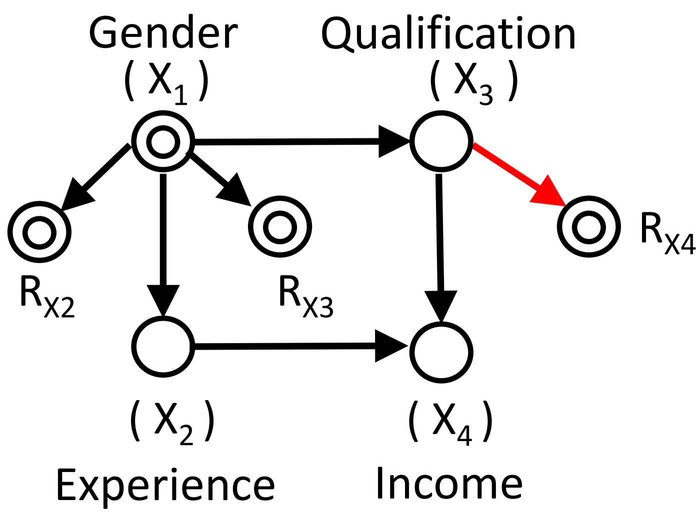
Informed Deletion

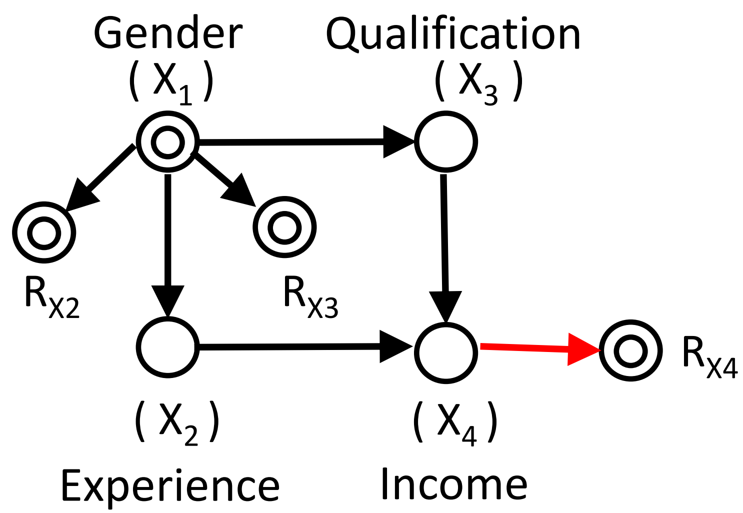
$$P(X_1, X_2) = \boldsymbol{P\left(X_2 | X_1, R_{X_2} = ob\right)} P(X_1)$$

Missing Not At Random (MNAR)

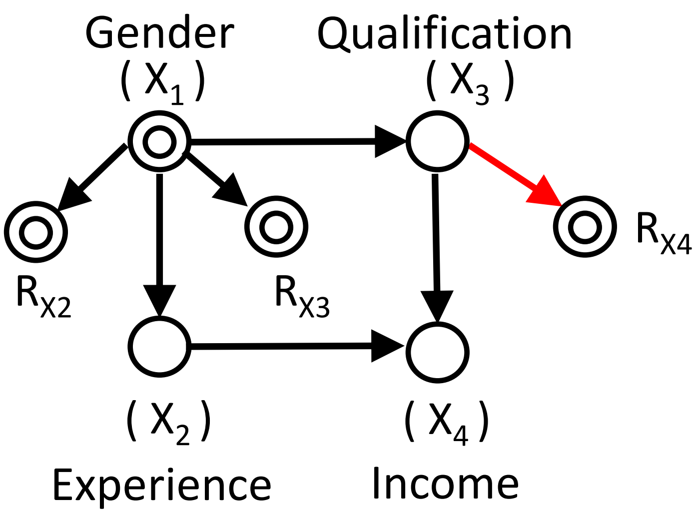


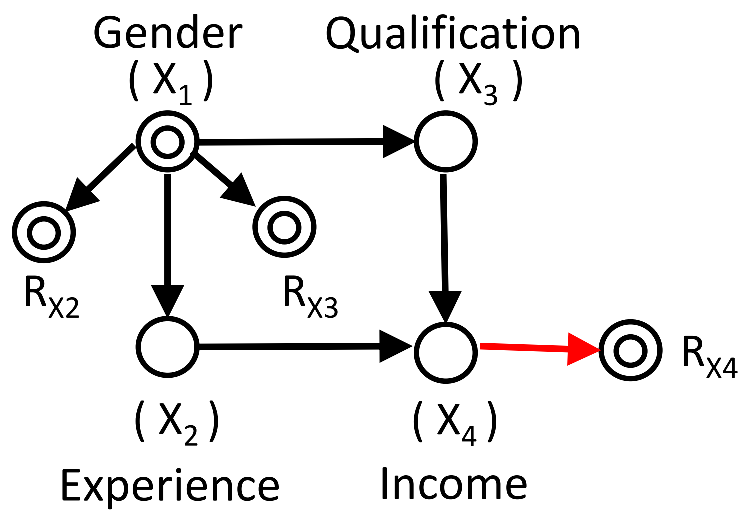
Missing Not At Random (MNAR)

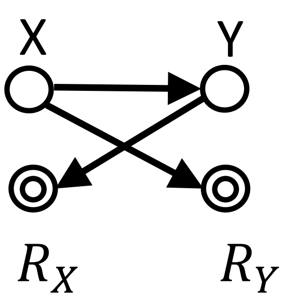




Missing Not At Random (MNAR)







$$P(X,Y) = \frac{P(R_X = ob, R_Y = ob, X, Y)}{P(R_X = ob | Y, R_Y = ob) P(R_Y = ob | X, R_X = ob)}$$

Conclusions

- Everybody loves to hate EM (slow, stuck, etc.)
- Deletion is solution to some EM problems
- Opens doors
  - Big incomplete data
  - Consistent learning of intractable networks
  - Efficient structure learning from incomplete data
  - Learning from MNAR data
- Surprising (given BN textbooks)?
- Code: http://reasoning.cs.ucla.edu/deletion

Thanks!