Model based and model assisted estimators using probabilistic expert systems

Paola Vicard (Università Roma Tre)

Joint work with:
Marco Ballin and Mauro Scanu
(ISTAT)

Durham, 2 July 2008

Let $\mathcal{P}$ be a finite population of size N.

Let $Y_1, \ldots, Y_k$ be k categorical variables of interest with distribution

Parameter of interest → $$\theta_{y_1,\ldots,y_k} = \sum_{i=1}^{N} \frac{I_{y_1 \ldots y_k}(y_{i1},\ldots,y_{ik})}{N}$$

$$I_{y_1 \ldots y_k}(y_{i1},\ldots,y_{ik}) = \begin{cases} 1, & \text{if } (y_{i1},\ldots,y_{ik}) = (y_1,\ldots,y_k) \\ 0 & \text{otherwise} \end{cases} \qquad i = 1,\ldots,N$$

We are interested in estimating a contingency table.

$\theta_{y_1,\ldots,y_k}$ can be a complex object (complexity being due to the number of variables, the number of variable categories, and the association structure among variables). ***The relation structure can help in finding an efficient estimator.***

Let $\mathcal{S}$ be a sample drawn from $\mathcal{P}$ according to a stratified sampling design with H strata $s_h, h = 1,\ldots,H$, and corresponding survey weights w_h.

The Horvitz-Thompson estimator of $\theta_{y_1,\ldots,y_k}$ is

$$\hat{\theta}_{y_1,\ldots,y_k} = \sum_{i\in\mathcal{S}} I_{y_1\ldots y_k}(y_{i1},\ldots,y_{ik})\frac{w_i}{\sum_{i\in\mathcal{S}} w_i} = \sum_{h=1}^{H}\frac{w_h}{N}\sum_{i\in s_h} I_{y_1\ldots y_k}(y_{i1},\ldots,y_{ik})$$

unit sampling weight

$w_i = w_h$ for $i \in s_h, h = 1,\ldots,H$

Here the design variables are merged to produce an *adequate summary* (in the sense of Rubin, 1985) that is a summary variable *SD* with as many states (*H*) as the strata.

$$\theta_h = \sum_{i\in\mathcal{S}}\frac{I_{w_h}(w_i)\,w_i}{\sum_{i\in\mathcal{S}} w_i} = \frac{n_h w_h}{\sum_{h=1}^{H} n_h w_h} \qquad h = 1,\ldots,H$$

If H is larger than the number of different inclusion probabilities then the weights can be defined as w_h/h, $h=1,\ldots, H$ (Smith T.M.F., 1988)

Aim of this work:

Exploit information on the multivariate dependency structure to propose a class of estimators for $\theta_{y_1,\ldots,y_k}$

Proposed tool:

Probabilistic Expert Systems (PES)

Why probabilistic expert systems?

Descriptive advantage (the dependence relationship among variables can be easily read from the graphical structure).

PES allows using easy and computationally efficient algorithms for evidence propagation.

↓

PES help updating multivariate distributions given auxiliary information (integration of different sources; coherence between estimates from different surveys)

Possibility to formalize post stratification via graphical models

PES are useful for evaluation of possible scenarios and for supporting *decision makers*

PES and sampling from finite population

Recall that *SD* is a categorical variable representing the stratified sampling design, *i.e.* with as many states as the strata

$$\theta_h = \frac{n_h w_h}{\sum_{h=1}^{H} n_h w_h} \qquad h = 1,\ldots,H$$

Conditionally on *SD,* the survey weights w_h are *hidden* in the estimation of the marginal and conditional distributions of the variables of interest

$$\hat{\theta}_{y_j|h} = \frac{\sum_{i\in s_h} I_{y_j}(y_{ij})w_h}{\sum_{i\in s_h} w_h} = \frac{\sum_{i\in s_h} I_{y_j}(y_{ij})}{n_h}$$

$$\hat{\theta}_{y_j|h,Y_l=y_l} = \frac{\sum_{i\in s_h} I_{y_j y_l}(y_{ij}, y_{il})}{\sum_{i\in s_h} I_{y_l}(y_{il})}$$

PES based estimators

Assume a *PES* for *SD*, $Y_1, \ldots, Y_k$ – *SD founder node*

The joint probability distribution of (SD, $Y_1, \ldots, Y_k$) is

$$\theta_{h,y_1,\ldots,y_k} = \theta_h \theta_{y_1|h} \theta_{y_2|h,y_1} \cdots\cdots \theta_{y_k|h,y_1,\ldots,y_{k-1}} = \theta_h \prod_{j=1}^{k} \theta_{y_j|pa(y_j)}$$

Therefore the ***PES based* estimator** (in a model based approach where the design variables are modelled together with the variables of interest) is

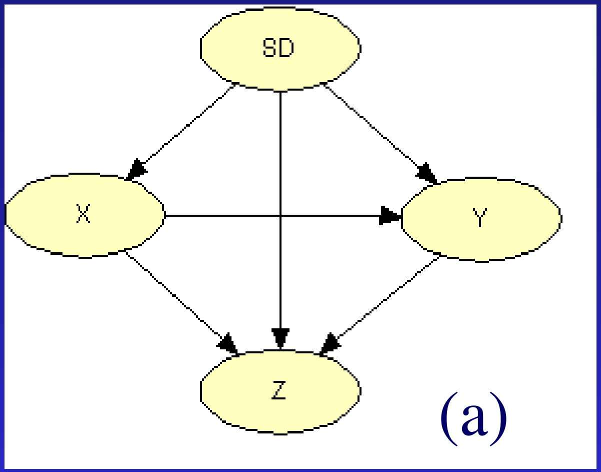
$$\hat{\theta}_{y_1,\ldots,y_k} = \sum_{h=1}^{H} \theta_h \hat{\theta}_{y_1|h} \hat{\theta}_{y_2|y_1,h} \cdots \hat{\theta}_{y_k|y_1,\ldots y_{k-1},h} = \sum_{h=1}^{H} \theta_h \prod_{j=1}^{k} \hat{\theta}_{y_j|pa(y_j)}$$

→ θ_h is not sample based because it is known by design.

Examples

Consider 3 variables of interest *X*, *Y*, *Z*

Suppose the *PES* is complete



Applying the *chain rule* to (*SD*, *X*, *Y*, *Z*) in model (a) we have

$$\theta_{h,x,y,z} = \theta_h \theta_{x|h} \theta_{y|x,h} \theta_{z|x,y,h}$$

Marginalizing with respect to *SD* the estimator based on the complete model is

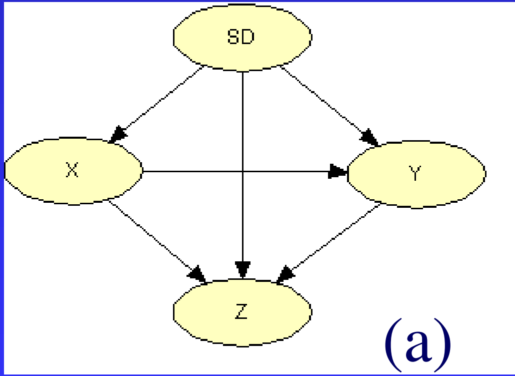
$$\hat{\theta}^{(a)}_{x,y,z} = \sum_{h=1}^{H} \theta_h \hat{\theta}_{x|h} \hat{\theta}_{y|x,h} \hat{\theta}_{z|x,y,h}$$

It can be shown that $\hat{\theta}^{(a)}_{x,y,z}$ coincides with the Horvitz-Thompson estimator

$$\hat{\theta}^{(a)}_{x,y,z} = \sum_{h=1}^{H} \theta_h \hat{\theta}_{x|h} \hat{\theta}_{y|x,h} \hat{\theta}_{z|x,y,h} =$$

$$= \sum_{h=1}^{H} \frac{n_h w_h}{\sum_{h=1}^{H} n_h w_h} \sum_{i \in s_h} \frac{I_x(x_i)}{n_h} \sum_{i \in s_h} \frac{I_{x,y}(x_i, y_i)}{\sum_{i \in s_h} I_x(x_i)} \sum_{i \in s_h} \frac{I_{x,y,z}(x_i, y_i, z_i)}{\sum_{i \in s_h} I_{x,y}(x_i, y_i)} =$$

$$= \sum_{h=1}^{H} \frac{w_h}{\sum_{h=1}^{H} n_h w_h} \sum_{i \in s_h} I_{x,y,z}(x_i, y_i, z_i)$$

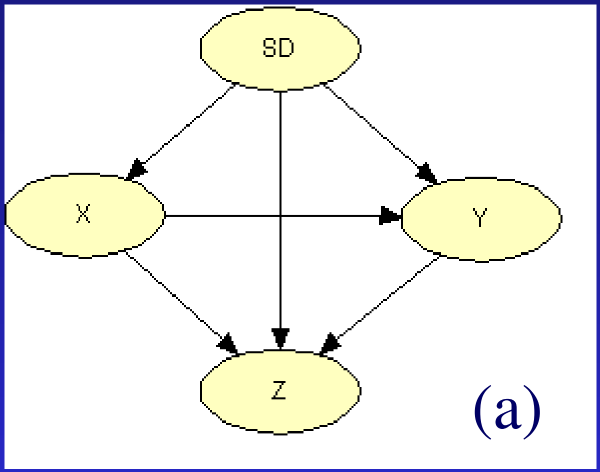


(a)

The Horvitz-Thompson estimator can be interpreted as a model based estimator relying on the **complete model**.

On the use of the complete graphical model

Problem: possible overparameterization

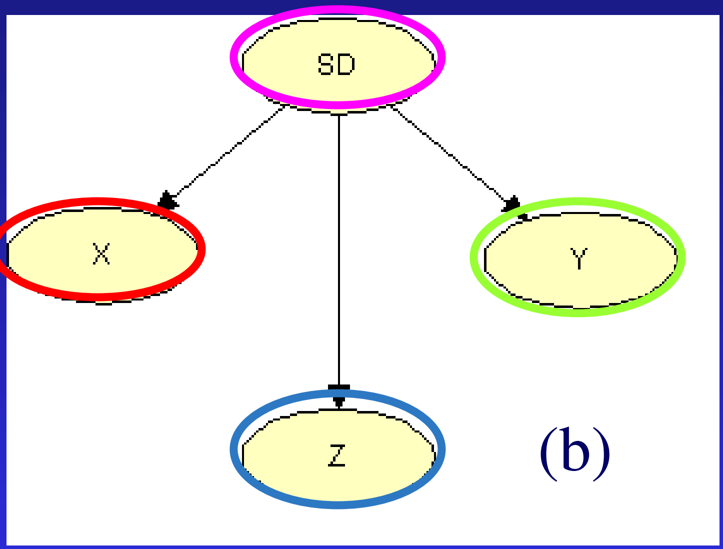


$\hat{\theta}^{(a)}_{x,y,z}$ could be less efficient than the estimator based on the actual association structure among the variables.

Proposed solution: given a *PES* structure, use the corresponding *PES* based estimator

$$\hat{\theta}^{(PES)}_{x,y,z} = \sum_{h=1}^{H} \theta_h \hat{\theta}_{x|pa(x)} \hat{\theta}_{y|pa(y)} \hat{\theta}_{z|pa(z)}$$

Examples of non complete models: 1

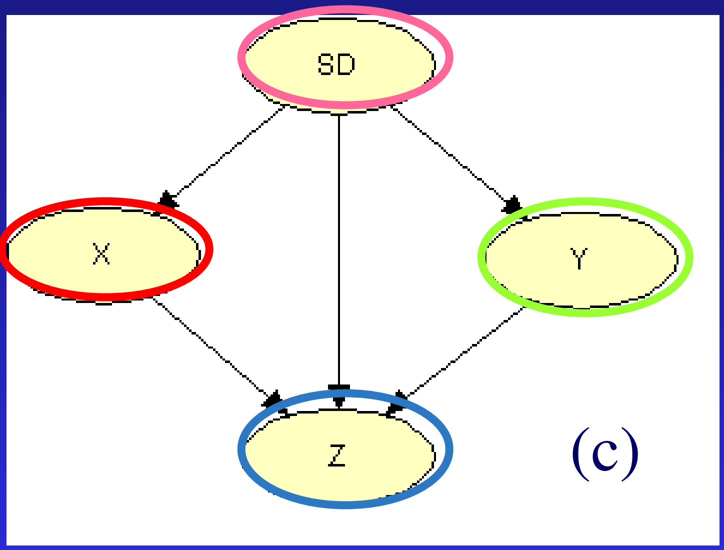


X, *Y* and *Z* are independent given *SD*.

$$\hat{\theta}^{(b)}_{x,y,z} = \sum_h \theta_h \hat{\theta}_{x|h} \hat{\theta}_{y|h} \hat{\theta}_{z|h}$$

$$= \sum_{h=1}^{H} \frac{n_h w_h}{\sum_h n_h w_h} \frac{\sum_{i \in s_h} I_x(x_i)}{n_h} \frac{\sum_{i \in s_h} I_y(y_i)}{n_h} \frac{\sum_{i \in s_h} I_z(z_i)}{n_h}$$

Examples of non complete models: 2

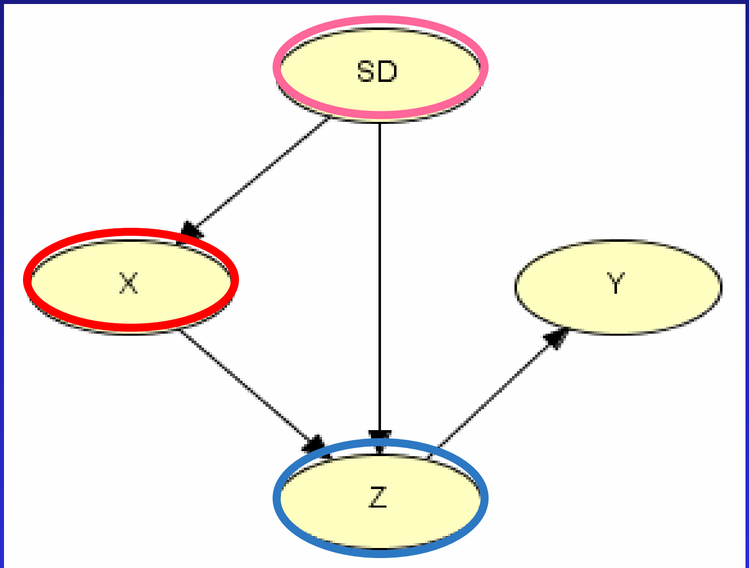


X and Y are independent given SD but dependent given Z.

$$\hat{\theta}^{(c)}_{x,y,z} = \sum_h \theta_h \hat{\theta}_{x|h} \hat{\theta}_{y|h} \hat{\theta}_{z|x,y,h}$$

$$= \sum_{h=1}^{H} \frac{n_h w_h}{\sum_{h=1}^{H} n_h w_h} \sum_{i \in s_h} \frac{I_x(x_i)}{n_h} \sum_{i \in s_h} \frac{I_y(y_i)}{n_h} \sum_{i \in s_h} \frac{I_{x,y,z}(x_i, y_i, z_i)}{\sum_{i \in s_h} I_{x,y}(x_i, y_i)}$$

Examples of non complete models: 3



There is no direct connection between *SD* and *Y*.

$$\hat{\theta}^{(d)}_{x,y,z} = \sum_h \theta_h \hat{\theta}_{x|h} \hat{\theta}_{z|x,h} \hat{\theta}_{y|z}$$

$$= \sum_{h=1}^{H} \frac{n_h w_{(h)}}{\sum_{h=1}^{H} n_h w_{(h)}} \sum_{i \in s_h} \frac{I_x(x_i)}{n_h} \sum_{i \in s_h} \frac{I_{x,z}(x_i, z_i)}{\sum_{i \in s_h} I_x(x_i)} \left(\sum_{i \in s} \frac{I_{y,z}(y_i, z_i)}{\sum_{i \in s_h} I_z(z_i)} \right)$$

Some considerations

$\hat{\theta}^{(a)}_{x,y,z}$, Horvitz-Thompson estimator, is consistent and unbiased

$\hat{\theta}^{(PES)}_{x,y,z}$ is consistent but not unbiased.

Concerning each factor $\hat{\theta}^{(PES)}_{y|pa(y)}$ in the chain rule.

$\hat{\theta}^{(PES)}_{y|pa(y)}$ has a smaller variance compared to factors with a larger parent set; hence there is a gain in terms of variance of $\hat{\theta}^{(PES)}_{y|pa(y)}$ with respect to $\hat{\theta}^{(a)}_{y|pa(y)}$

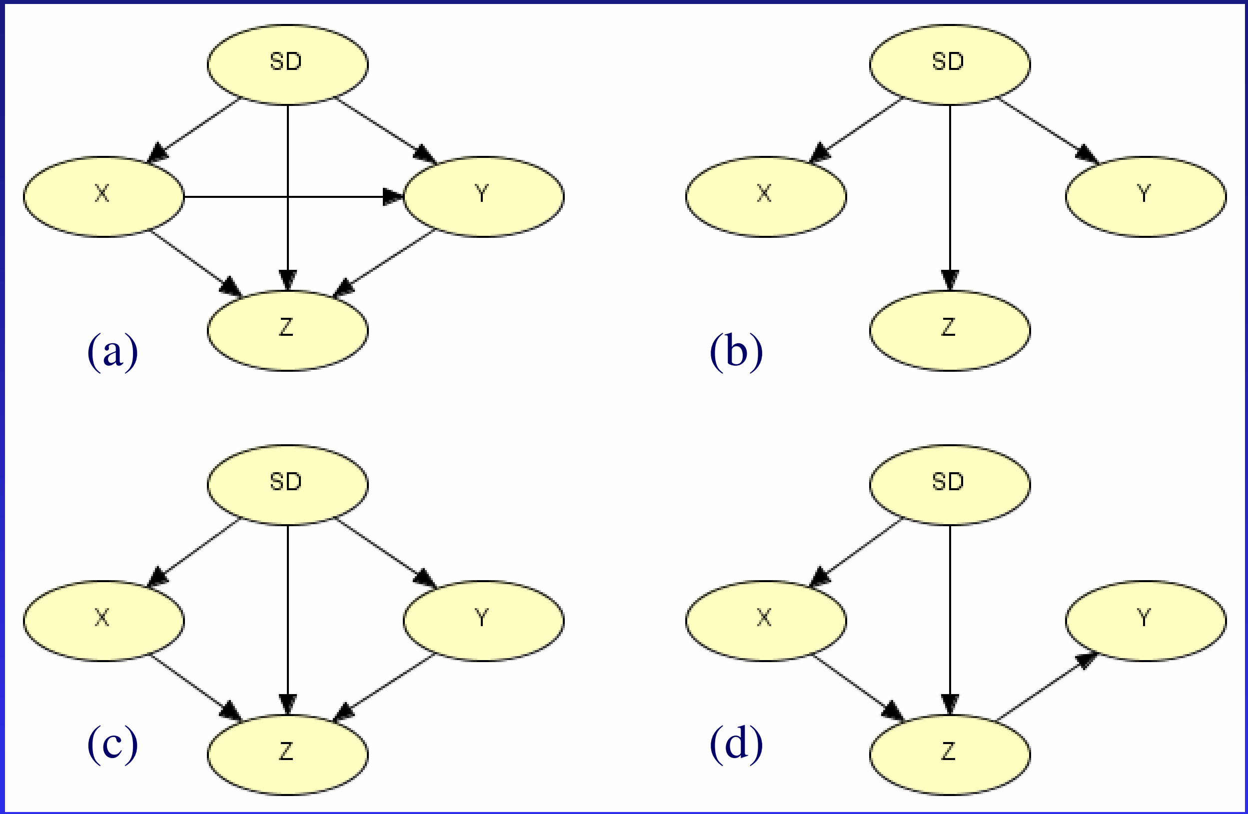
$\hat{\theta}^{(PES)}_{y|pa(y)}$ is less biased compared to factors with a smaller parent set.

The lack of *true* parents effect is predominant

Monte Carlo experiment

4 populations with 10000 units have been generated according to 4 structures.

- *X* (2 categories)
- *Y* (3 categories)
- *Z* (2 categories)



From each population 1000 samples of size n=1000 have been drawn according to a stratified sampling design with 3 strata.

- *SD* has 3 categories

Monte Carlo experiment

Stratum code h	Stratum size N_h	θ_h	Sample size n_h
$h=1$	5995	0,5995	100
$h=2$	2959	0,2959	200
$h=3$	1046	0,1046	700

Note that the sampling fraction is not proportional to stratum size

The performances of the different estimators are measured and compared by the Monte Carlo estimates of the chi-square distance between the two joint distributions:

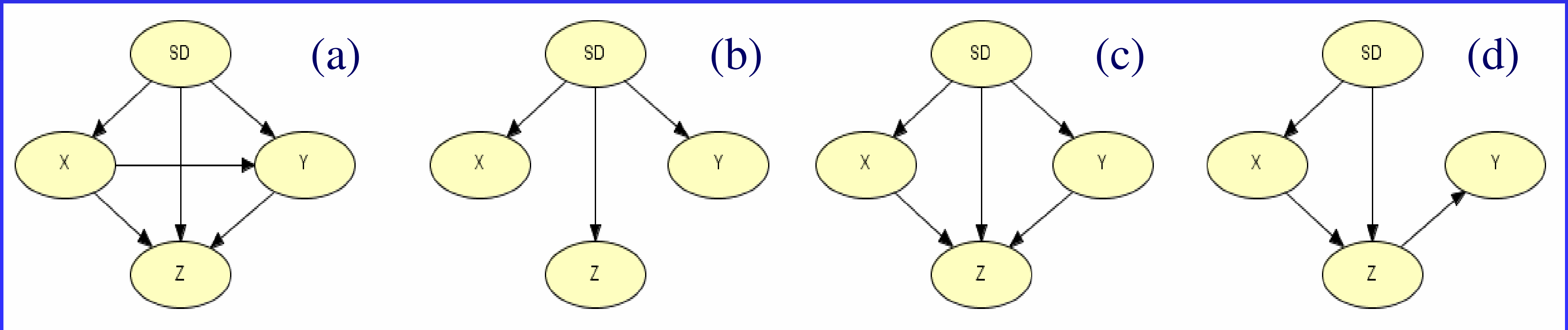
$$\chi\left(\hat{\theta}^{(PES)}_{x,y,z}\right) = \frac{1}{M}\sum_{m=1}^{M}\sum_{x,y,z}\frac{\left[\hat{\theta}^{(PES),m}_{x,y,z} - \theta_{x,y,z}\right]^2}{\theta_{x,y,z}}$$

M=1000=number of Montecarlo replications

Monte Carlo experiment

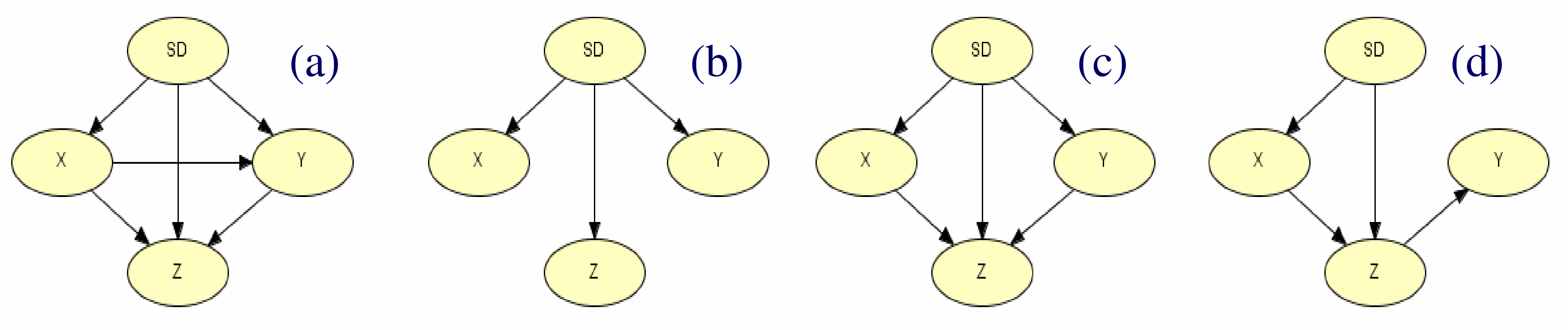
Pop	$\chi\left(\hat{\theta}^{(a)}_{x,y,z}\right)$	$\chi\left(\hat{\theta}^{(b)}_{x,y,z}\right)$	$\chi\left(\hat{\theta}^{(c)}_{x,y,z}\right)$	$\chi\left(\hat{\theta}^{(d)}_{x,y,z}\right)$
a	37.5	64.4	40.7	377.6
b	30.5	17.9	26.0	382.8
c	32.7	51.6	28.9	1227.2
d	34.6	32.6	29.3	13.2

Estimator based on (d) seems less robust than those based on (a) - (c)



Monte Carlo experiment

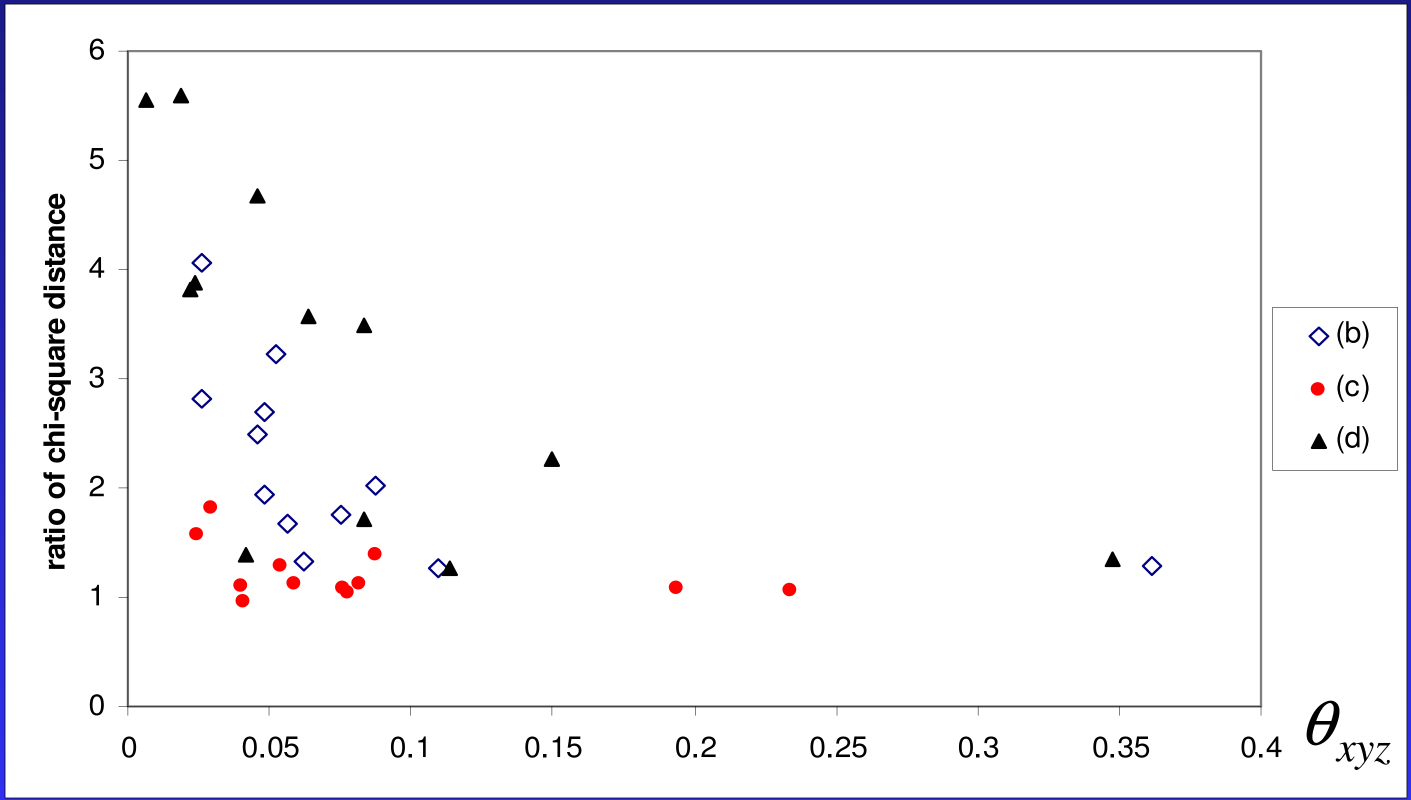
Pop	$Bias_{(a)}$	$Bias_{(b)}$	$Bias_{(c)}$	$Bias_{(d)}$
a	0.04	73.8	13.7	96.3
b	0.09	2.09	0.91	96.1
c	0.10	71.3	1.21	98.5
d	0.06	46.5	0.92	1.4



Estimators based on the correct model structure are approximately unbiased.

Monte Carlo experiment

(probability estimates of each single cell)



$$\frac{\chi\left(\hat{\theta}^{(a)}_{x,y,z}\right)}{\chi\left(\hat{\theta}^{(PES)}_{x,y,z}\right)}$$

Ratio of the Monte Carlo estimates of the chi-square distance of the *PES*-estimators based on the correct structure

Problem:

If based on a structure where one or more variables of interest are not children of the sampling design node *SD*, *PES*-based estimators are not robust to model miss-specification.

A possible solution?

Definition of estimators in a *model assisted* framework

- The design variable *SD* is not directly modelled with the variables of interest
- Information on design variables is incorporated via survey weights

PES assisted estimators

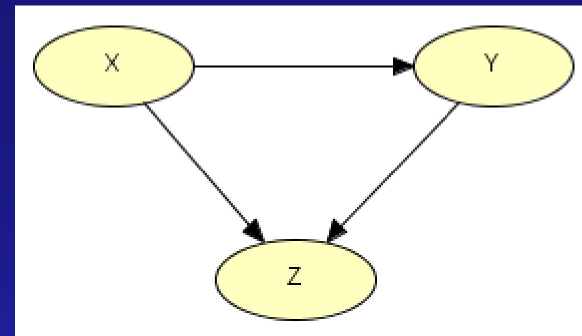
Consider a *PES* for $(Y_1,\ldots, Y_k)$ with $\theta_{y_1,\ldots,y_k} = \prod_{j=1}^{k} \theta_{y_j|pa(y_j)}$

The *PES assisted* estimator is $\hat{\hat{\theta}}_{y_1,\ldots,y_k} = \prod_{j=1}^{k} \hat{\hat{\theta}}_{y_j|pa(y_j)}$

Where each factor is a weighted estimator of the conditional distributions

$$\hat{\hat{\theta}}_{y_j|pa(y_j)} = \sum_{i=1}^{n} \frac{w_i I_{y_j,pa(y_j)}\left(y_{ij}, pa\left(y_{ij}\right)\right)}{\sum_{i=1}^{n} w_i I_{pa(y_j)}\left(pa\left(y_{ij}\right)\right)}$$

Example: the complete graph



$$\hat{\hat{\theta}}_{x,y,z} = \hat{\hat{\theta}}_{x}\hat{\hat{\theta}}_{y|x}\hat{\hat{\theta}}_{z|x,y} =$$

$$= \sum_{i=1}^{n} \frac{w_i I_x(x_i)}{\sum_{i=1}^{n} w_i} \sum_{i=1}^{n} \frac{w_i I_{x,y}(x_i, y_i)}{\sum_{i=1}^{n} w_i I_x(x_i)} \sum_{i=1}^{n} \frac{w_i I_{x,y,z}(x_i, y_i, z_i)}{\sum_{i=1}^{n} w_i I_{x,y}(x_i, y_i)} =$$

$$= \sum_{i=1}^{n} \frac{w_i I_{x,y,z}(x_i, y_i, z_i)}{\sum_{i=1}^{n} w_i} =$$

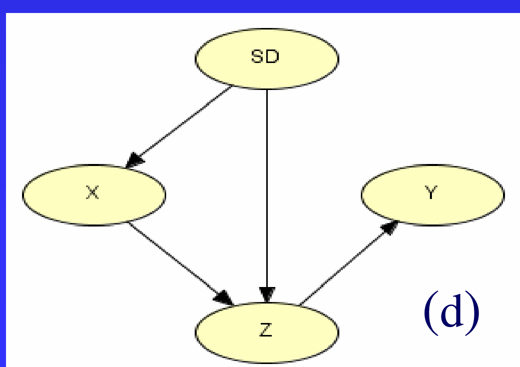
The *PES assisted* estimator referring to the complete model coincides with the Hotviz-Thompson estimator.

$$= \hat{\theta}^{(a)}_{x,y,z}$$

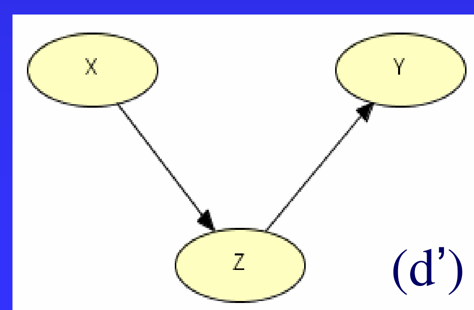
The complete model is the only PES whose corresponding *model based* and *model assisted* estimators are “compatible”

Monte Carlo experiment

Pop	$\chi\left(\hat{\theta}^{(d)}_{x,y,z}\right)$	$\chi\left(\hat{\hat{\theta}}^{(d')}_{x,y,z}\right)$
a	377.6	60.9
b	382.8	49.1
c	1227.2	133.9
d	13.2	25.3



(d)

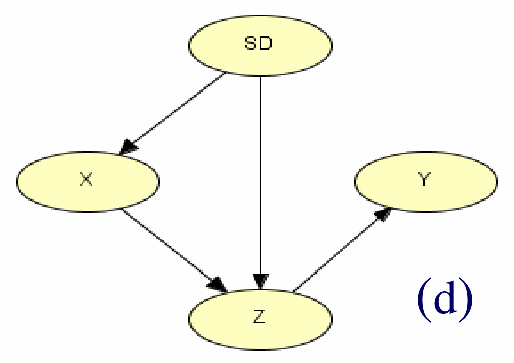


(d’)

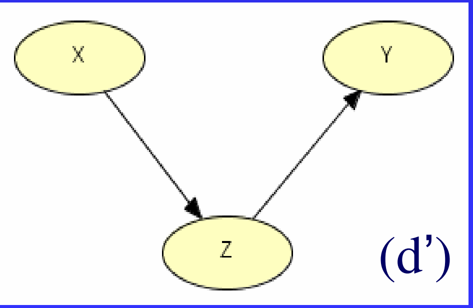
$$\hat{\theta}^{(d')}_{x,y,z} = \hat{\hat{\theta}}_x \hat{\hat{\theta}}_{y|z} \hat{\hat{\theta}}_{z|x}$$

Monte Carlo experiment

Pop	$\text{Bias}\left(\hat{\theta}^{(d)}_{x,y,z}\right)$	$\text{Bias}\left(\hat{\hat{\theta}}^{(d')}_{x,y,z}\right)$
a	96.3	58.1
b	96.1	57.5
c	98.5	83.5
d	1.4	1.4



(d)



(d')

$$\hat{\theta}^{(d')}_{x,y,z} = \hat{\hat{\theta}}_x \hat{\hat{\theta}}_{y|z} \hat{\hat{\theta}}_{z|x}$$

Structural learning
(maximum likelihood structural learning)

Given a *PES* for (SD, $Y_1,\ldots,Y_k$) – *SD root*, the joint probability distribution is

$$\theta_{h,y_1,\ldots,y_k} = \theta_h \prod_{j=1}^{k} \theta_{y_j|pa(y_j)}$$

Given a *PES*, the likelihood on the sample is

$$L\left(\theta_{hy_1,\ldots,y_k}; PES\right) = \prod_{i=1}^{n} \theta_h^{w_i} \prod_{j=1}^{k} \theta_{y_j|pa(y_j)}^{I(y_j|pa(y_j))}$$

The maximum likelihood estimator of the parameters is the *PES* based estimator

To estimate the structure we consider the likelihood as a function of *PES*; the penalised loglikelihood function

$$s(PES) = \log L\left(\hat{\theta}_{hy_1\ldots y_k}; PES\right) - \frac{\log n}{2} Q$$

Number of parameters in the model

The best *PES* is that with the highest score

Propagation and Poststratification

Suppose an informative shock occurs to variable *X* whose updated frequency distribution is

$$N^*_{x_q}, \quad q = 1, \ldots, Q$$

Q = n° of states of variable X

By propagating this information through the network, we poststratify the sample with respect to *X*.

The original sample weights w_i are updated so that the estimators verify the new constraints on *X*.

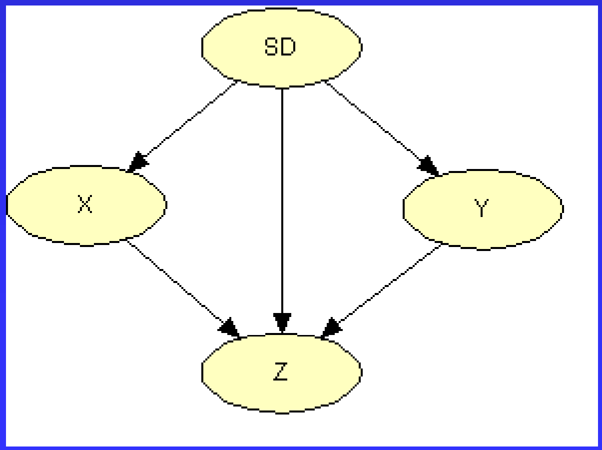
$$w_i^* = w_i \frac{N^*_{x_q}}{\sum_i w_i I_{x_i}(q)} = w_i \frac{N^*_{x_q}}{\hat{N}_{x_q}}, \quad i : I_{x_i}(q) = 1, \quad q = 1, \ldots, Q$$

update ratio

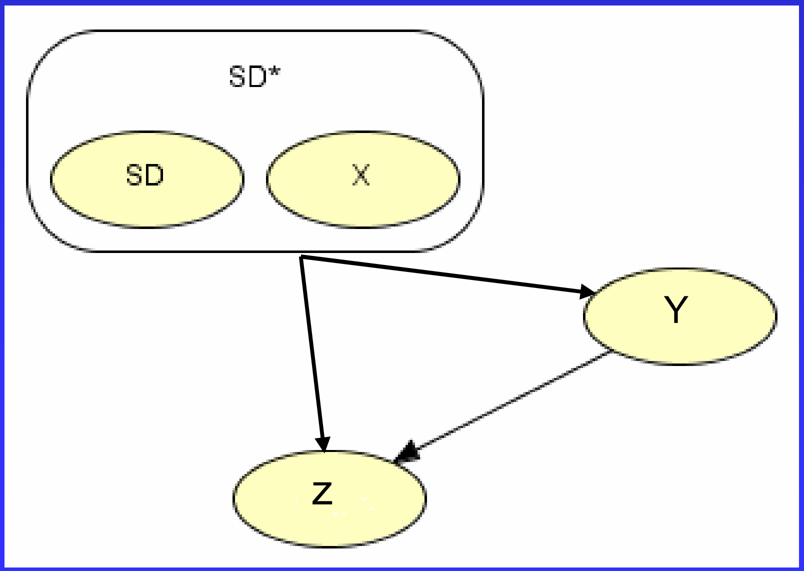
Poststratification

From a *graphical* point of view, poststratification corresponds to modify node *SD* into a new node *SD** such that:

- *SD** strata are given by the Cartesian product of *SD* and *X* categories, *i.e.* (h, q), $h = 1,\ldots,H$, $q = 1, \ldots,Q$ $w^{*}_{(h,q)}$
- The units in the same category (h, q) have the same weight



Poststratification with respect to *X*



Poststratification
(weights computation)

By poststratification we update the joint distribution θ_{h,x_q}

$$\theta^*_{h,x_q} = \theta_{h|x_q} \boxed{\theta^*_{x_q}} =$$

$\theta^*_{x_q}$ new frequency of category x_q of X

$$= \frac{\theta_h \theta_{x_q|h}}{\sum_{h=1}^{H} \theta_h \theta_{x_q|h}} \theta^*_{x_q} = \frac{n_h w_h}{\sum_{h=1}^{H} n_h w_h} \frac{n_{hq}}{n_h} \frac{\theta^*_{x_q}}{\hat{\theta}_{x_q}}, \; q = 1, \ldots, Q \text{ e } h = 1, \ldots, H$$

Units in the same category (h, q) of SD^* have the same weight. Let n_{hq} be the size of (h, q), hence

$$w^*_{(h,q)} = \frac{\sum_{h=1}^{H} n_h w_h}{n_{hq}} \theta^*_{h,x_q} = w_h \frac{\theta^*_{x_q}}{\hat{\theta}_{x_q}}, \; q = 1, \ldots, Q \text{ e } h = 1, \ldots, H$$

PES structures for *model assisted* estimators