A First Inquiry into Simpson's Paradox with Belief Functions

François Delmotte, David Mercier, and Frédéric Pichon

Univ. Lille Nord de France, F-59000 Lille, France
UArtois, LGI2A, F-62400, Béthune, France
firstname.lastname@univ-artois.fr

Abstract. Simpson's paradox, also known as the Yule-Simpson effect, is a statistical paradox which plays a major role in causality modelling and decision making. It may appear when marginalizing data: an effect can be positive for all subgroups of a population, but it can be negative when aggregating all the subgroups. This paper explores what happens if data are considered in the framework of belief functions instead of classical probability theory. In particular, the co-occurrence of the paradox with both the probabilistic approach and our belief function approach is studied.

Keywords: Simpson's paradox, Belief functions, Marginalization, Decision Making.

1 Introduction

First relations about Simpson's paradox, also known as the Yule-Simpson effect or the reversal paradox, were discovered one century ago [12], and more studies were published from the fifties [11]. It concerns statistical results that are very strange to the common sense: the influence of one variable can be positive for every subgroups of a population, but it may be negative for the whole population. For instance a medical treatment can be effective when considering the gender of patients, being effective for both males and females separately. But when studying the population as a whole, when the gender is ignored by marginalizing this variable, the treatment becomes inefficient. Real life examples are numerous. In [12] several examples are given concerning batting averages in baseball, kidney stone treatment, or the Berkeley sex bias case when studying the admittance at this university depending on the gender of the applicants.

Simpson's paradox plays a major role in causality modelling, when it is ignored if a variable has an influence on the others, or not. Since it has been shown that given a set of data involving several variables, any relation may be reversed when marginalizing (see [7] for a graphical proof), Simpson's paradox has a direct role in the adjustment problem, when somebody tries to know what variables have to be taken into account in order to obtain a model of the process. A proposed solution to this paradox does not lie in the original data, but involves extraneous information, provided by experts [9], called causal relations.

A. Laurent et al. (Eds.): IPMU 2014, Part III, CCIS 444, pp. 190–199, 2014.

So far Simpson's paradox phenomenon has been described with probability measures. However, in the last 40 years, other models of uncertainty have been defined, and in particular belief functions. Introduced by Shafer [13], they have wider and wider domains of application (see in particular [10, Appendix A]).

It would be interesting to know if such reversal of the decisions also occurs with belief functions when marginalizations are involved. Through simulations, a first inquiry into that question is presented in this paper.

This paper is organized as follows. A few details about the paradox are given in Section 2 and basic concepts on belief functions are exposed in Section 3. Then, in Section 4 is presented a belief function approach to handle data at the origin of the paradox with an academic example. Then a Monte Carlo experiment is provided in Section 5 to provide the (co) occurences of the paradox. Section 6 concludes this paper.

2 Simpson's Paradox

Introducing Simpson's paradox may be done purely arithmetically with 8 numbers such that:

$$a/b < A/B \tag{1}$$

$$c/d < C/D \tag{2}$$

$$(a+c)/(b+d) <> (A+C)/(B+D) \tag{3}$$

In the last equation (Eq. 3), it is not known what terms is the biggest. In other words, (Eq. 1) and (Eq. 2) are insufficient to deduce any order in (Eq. 3).

For example, the next relations [12] hold:

$$\frac{1}{5} < \frac{2}{8} \text{ (Eq. 1)}, \ \frac{6}{8} < \frac{4}{5} \text{ (Eq. 2), and } \frac{7}{13} > \frac{6}{13} \text{ (Eq. 3).} \tag{4}$$

When considering probabilities, these sets of inequalities can be seen as:

$$P(A \mid B, C) < P(A \mid \overline{B}, C) \tag{5}$$

$$P(A \mid B, \overline{C}) < P(A \mid \overline{B}, \overline{C}) \tag{6}$$

$$P(A \mid B) <> P(A \mid \overline{B}) \tag{7}$$

This is where Simpson's paradox appears [9]: a property may be true in every subgroups of a population while being false for the whole population (In the previous example two subgroups corresponding to C and $\overline{C}$ are considered).

Real life examples of such discrepancies are numerous. They are even encountered in evolutionary games involving populations of rats and lemmings [2,7].

Such an example [9,12] about a medical treatment is presented in Table 1. It can be observed that the treatment is effective when the variable Gender is taken into account. Indeed:

$$P(S \mid A, m) \simeq .93 > P(S \mid B, m) \simeq .87 \; , \tag{8}$$

$$P(S \mid A, f) \simeq .73 > P(S \mid B, f) \simeq .69 \; . \tag{9}$$

Table 1. Probabilities of success (S) and failure (F) of a treatment knowing it has been given (A) or not (B) when a male m is encountered versus a female f, a number of 700 cases being considered

	m		f	
	A	B	A	B
S	$81/87 \simeq 0.93$	$234/270 \simeq 0.87$	$192/263 \simeq 0.73$	$55/80 \simeq 0.69$
F	$6/87 \simeq 0.07$	$36/270 \simeq 0.13$	$71/263 \simeq 0.27$	$25/80 \simeq 0.31$

However, when the gender is marginalized, for the whole population, no treatment becomes a better option. Indeed, as it can be seen in Table 2, inequalities are reversed:

$$P(S \mid A) \simeq .78 < P(S \mid B) \simeq .83 \; . \tag{10}$$

Table 2. Probability of success and failure obtained from Table 2 when the gender is marginalized

	A (with treatment)	B (no treatment)
S	$P_{11} = 273/350 \simeq 0.78$	$P_{12} = 289/350 \simeq 0.83$
F	$P_{21} = 77/350 \simeq 0.22$	$P_{22} = 61/350 \simeq 0.17$

Let us also note that changing numbers while keeping ratios constant, with $a/b = (\alpha a)/(\alpha b)$ and α free, may alter the paradox. Indeed ratios in (Eq. 1) and (Eq. 2) are constant, as is the right term of (Eq. 3), but the left term of the latter now depends on α. For example consider $\alpha = 10$, then equation 4 becomes $10/50 < 2/8$, $6/8 < 4/5$ and $16/58 < 6/13$, *i.e*, the opposite conclusion is reached.

A solution proposed by Pearl [9] consists in using extraneous data, called causal information, that are different from the raw data. Causal information is the knowledge of the influences of some variables on some other ones. They enable one to know if one must reason with the full contingency table, or the reduced one. So the conclusions drawn from each table cannot be compared, and so the paradox becomes impossible. However, causal information must be provided by experts and therefore may be unavailable or difficult to obtain.

3 Belief Functions

This section recalls some basic concepts on belief functions which are used in this paper.

3.1 Main Functions

Let $\Omega = \{\omega_1, \ldots, \omega_n\}$ be a finite set called the *frame of discernment*. The quantity $m(A) \in [0,1]$ with $A \subseteq \Omega$ is the part of belief supporting A that, due to a lack of information, cannot be given to any strict subset of A [14]. Mass function m (or basic belief assignment) has to satisfy:

$$\sum_{A \subseteq \Omega} m^{\Omega}(A) = 1 \ . \tag{11}$$

Throughout this article, 2^{Ω} represents all the subsets of Ω. Plausibility $Pl^{\Omega}(A)$ represents the total amount of belief that may be given to A with further pieces of evidence:

$$Pl^{\Omega}(A) = \sum_{A \cap B \neq \emptyset} m^{\Omega}(B) \ . \tag{12}$$

These functions are in one-to-one correspondence [13], so they are used indifferently with the same term ***belief function***. A set A such that $m(A) > 0$ is called a *focal element* of m. The vacuous belief function is defined by $m(\Omega) = 1$. It represents the total lack of information.

3.2 Refinement, Vacuous Extension and Marginalization

Let R be a mapping from 2^{Θ} to 2^{Ω} such that every singleton $\{\theta\}$, with $\theta \in \Theta$ is mapped into one or several elements of Ω, and such that all images $R(\{\theta\}) \subseteq \Omega$ form a partition of Ω. Such a mapping R is called a *refining*, Ω is called a *refinement* of Θ and Θ is called a *coarsening* of Ω [13].

Any belief function m^{Θ} defined on Θ can be extended to Ω. This operation is called the vacuous extension of m^{Θ} on Ω, it is denoted by $m^{\Theta\uparrow\Omega}$ and is defined by:

$$m^{\Theta\uparrow\Omega}(A) = \begin{cases} m^{\Theta}(B) & \text{if } A = R(B) \\ 0 & \text{otherwise} \end{cases} \tag{13}$$

The coarsening operation is the opposite step. Starting from a belief function defined on Ω, a belief function on Θ is defined trough a mapping R defined as previously. The problem is that R is generally not an onto mapping: usually there will be some focal elements of m^{Ω} that are not the images of sets of Θ by R. In this case Shafer [13, chapter 6, page 117] has introduced two envelopes, called *inner* and *outer reductions*.

The inner reduction (or lower envelop) $\underline{\Theta}$ and the outer reduction (or upper envelop) $\overline{\Theta}$ are mappings respectively defined, from 2^{Ω} to 2^{Θ}, for all $A \subseteq \Omega$ by:

$$\underline{\Theta}(A) = \{\theta \in \Theta,\ R(\{\theta\}) \subseteq A\} \ , \tag{14}$$

and

$$\overline{\Theta}(A) = \{\theta \in \Theta,\ R(\{\theta\}) \cap A \neq \emptyset\} \ . \tag{15}$$

The *inner reduction* and *outer reduction* on Θ of a belief function m^{Ω} defined on Ω are then respectively given for all $B \subseteq \Theta$ by:

$$\underline{m}^{\Theta}(B) = \sum_{A \subseteq \Omega,\ \underline{\Theta}(A)=B} m^{\Omega}(A) \ , \tag{16}$$

and

$$\overline{m}^{\Theta}(B) = \sum_{A \subseteq \Omega,\ \overline{\Theta}(A)=B} m^{\Omega}(A) \ . \tag{17}$$

Conceptually, *marginalization* is a special case of coarsening where $\Omega = X \times Y$ and either $\Theta = X$ or $\Theta = Y$ [13].

3.3 Pignistic Transformation

Decisions from belief functions can be made using the pignistic transformation $BetP$ justified in [14,15] based on rationality requirements and axioms. It is defined, for all ω of Ω, by:

$$BetP^{\Omega}(\{\omega\}) = \sum_{\omega \in A} \frac{m^{\Omega}(A)}{\mid A \mid (1 - m(\emptyset))} \ . \tag{18}$$

The converse operation of the pignistic rule has also been defined. It actually yields the least specific mass function whose pignistic transformation is the probability measure P^{Ω} [6]. Let $m^{\Omega}_{LSP(P^{\Omega})}$ denote this mass function. It has n nested focal elements :

$$m^{\Omega}_{LSP(P^{\Omega})}(\{\omega_i, \ldots, \omega_n\}) = (n - i + 1)(p_i - p_{i-1}), \quad \forall i \in \{1, \ldots, n\}, \tag{19}$$

with $0 < P^{\Omega}(\{\omega_1\}) = p_1 < \ldots < p_n$ and $p_0 = 0$. If some p_i are equal, the result is the same whatever the order.

4 Inferring Decisions from Contingency Tables with Belief Functions

4.1 Proposed Approach Outline

The whole approach used in this article is summarized in Figure 1.

From initial data on a 3 binary variables contingency table $\Omega = X \times Y \times Z$, a probability measure $\hat{P}^{\Omega}$ is estimated. From the same data a belief function

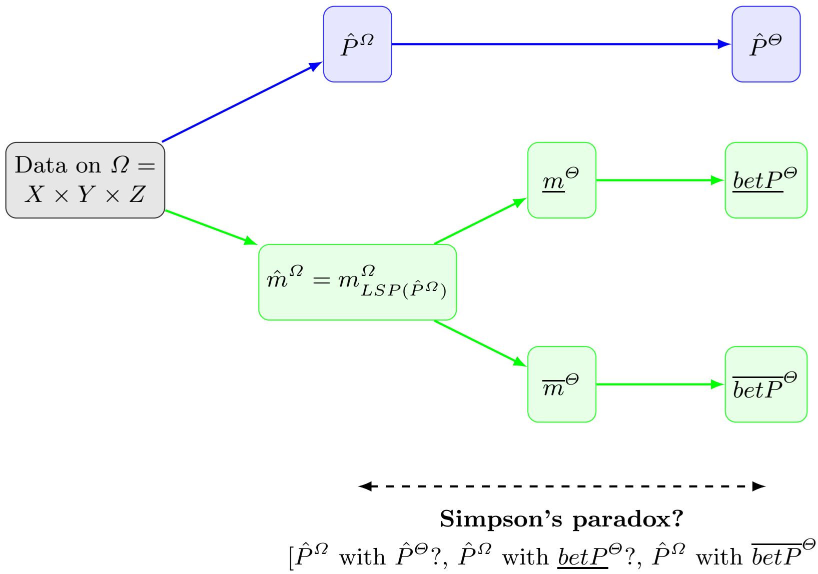

Fig. 1. Comparison of the Simpson's paradox when using probabilities and the proposed approach based on belief functions

$\hat{m}^{\Omega} = m^{\Omega}_{LSP(\hat{P}^{\Omega})}$ is estimated as the least specific belief function whose pignistic transformation gives $\hat{P}^{\Omega}$. It is obtained using Equation 19.

Then the probability measure $\hat{P}^{\Omega}$ and the belief function $\hat{m}^{\Omega}$ are marginalized on $\Theta = X \times Y$.

With belief functions, two reductions for this marginalization are considered: the inner reduction $\underline{m}^{\Theta}$ of $\hat{m}^{\Omega}$ (Eq. 16) and the outer reduction $\overline{m}^{\Theta}$ of $\hat{m}^{\Omega}$ (Eq. 17).

Finally, decisions based on the full contingency table are compared to decisions taken on the reduced space Θ to detect Simpson's paradoxes.

4.2 An Academic Example

In this section, an example is given to illustrate the approach with the inner reduction. Values in Tables 1 and 2 are used again.

Let us consider the spaces $\Omega = \{S, F\} \times \{A, B\} \times \{m, f\}$ and $\Theta = \{S, F\} \times \{A, B\}$ when the variable gender is marginalized.

In Table 3 are sorted out the numbers of cases of Table 1 to compute the least specific belief isopignistic to $\hat{P}^{\Omega}$.

Table 3. Computing $\hat{P}^{\Omega}$ from Table 1

Singleton	$\hat{P}^{\Omega}$
$FAm = (F, A, m)$	$\frac{6}{700} \simeq .01 = p_1$
FBf	$\frac{25}{700} \simeq .04 = p_2$
FBm	$\frac{36}{700} \simeq .05 = p_3$
SBf	$\frac{55}{700} \simeq .08 = p_4$
FAf	$\frac{71}{700} \simeq .10 = p_5$
SAm	$\frac{81}{700} \simeq .12 = p_6$
SAf	$\frac{192}{700} \simeq .27 = p_7$
SBm	$\frac{234}{700} \simeq .33 = p_8$

Using Equation 19, the least specific belief $\hat{m}^{\Omega}$ obtained from Table 3 can be computed:

$$\begin{aligned}
m^{\Omega}_{LSP(\hat{P})}(\Omega) &= 8p_1 &&= x_1\\
m^{\Omega}_{LSP(\hat{P})}(\Omega \setminus \{FAm\}) &= 7(p_2 - p_1) &&= x_2\\
m^{\Omega}_{LSP(\hat{P})}(\Omega \setminus \{FBf, FAm\}) &= 6(p_3 - p_2) &&= x_3\\
m^{\Omega}_{LSP(\hat{P})}(\Omega \setminus \{FBm, FBf, FAm\}) &= 5(p_4 - p_3) &&= x_4\\
m^{\Omega}_{LSP(\hat{P})}(\{SBm, SAf, SAm, FAf\}) &= 4(p_5 - p_4) &&= x_5\\
m^{\Omega}_{LSP(\hat{P})}(\{SBm, SAf, SAm\}) &= 3(p_6 - p_5) &&= x_6\\
m^{\Omega}_{LSP(\hat{P})}(\{SBm, SAf\}) &= 2(p_7 - p_6) &&= x_7\\
m^{\Omega}_{LSP(\hat{P})}(\{SBm\}) &= p_8 - p_7 &&= x_8
\end{aligned} \tag{20}$$

Mapping $\underline{\Theta}$ (Eq. 14) of $\Theta = \{S, F\} \times \{A, B\}$ is defined by:

$$\begin{aligned}
\underline{\Theta}(\Omega) &= \{SA, SB, FA, FB\}\\
\underline{\Theta}(\Omega \setminus FAm) &= \{SA, SB, FB\}\\
\underline{\Theta}(\Omega \setminus \{FBf, FAm\}) &= \{SA, SB\}\\
\underline{\Theta}(\Omega \setminus \{FBm, FBf, FAm\}) &= \{SA, SB\}\\
\underline{\Theta}(\{SBm, SAf, SAm, FAf\}) &= \{SA\}\\
\underline{\Theta}(\{SBm, SAf, SAm\}) &= \{SA\}\\
\underline{\Theta}(\{SBm, SAf\}) &= \emptyset\\
\underline{\Theta}(\{SBm\}) &= \emptyset
\end{aligned} \tag{21}$$

So the inner reduction $\underline{m}^{\Theta}$ of $\hat{m}^{\Omega}$ on the space Θ is given by (Eq. 16):

$$\underline{m}^{\Theta} : \begin{cases} \{SA, SB, FA, FB\} & \mapsto x_1 \\ \{SA, SB, FB\} & \mapsto x_2 \\ \{SA, SB\} & \mapsto x_3 + x_4 \\ \{SA\} & \mapsto x_5 + x_6 \\ \emptyset & \mapsto x_7 + x_8 \end{cases} \tag{22}$$

and pignistic values are the followings (Eq. 18):

$$\begin{array}{ll} \underline{BetP}^{\Theta}(\{SA\}) = k(\frac{x_1}{4} + \frac{x_2}{3} + \frac{x_3+x_4}{2} + x_5 + x_6) & \simeq 0.51 \\ \underline{BetP}^{\Theta}(\{SB\}) = k(\frac{x_1}{4} + \frac{x_2}{3} + \frac{x_3+x_4}{2}) & \simeq 0.31 \\ \underline{BetP}^{\Theta}(\{FA\}) = k(\frac{x_1}{4}) & \simeq 0.03 \\ \underline{BetP}^{\Theta}(\{FB\}) = k(\frac{x_1}{4} + \frac{x_2}{3}) & \simeq 0.15 \ , \end{array} \tag{23}$$

with $k = \frac{1}{1-x_7-x_8}$.

So $\underline{BetP}(S \mid A) \simeq \frac{.51}{.51+.03} \simeq 0.95$ and $\underline{BetP}(S \mid B) \simeq 0.68$, and unlike the probability case (Eq. 10) $\underline{BetP}(S \mid A) > \underline{BetP}(S \mid B)$, which leads to the same decision as the one obtained on the whole space $\{S, F\} \times \{A, B\} \times \{m, f\}$ (Eq. 8 and 9).

In this example, there is no longer a Simpson's paradox when considering belief functions and the inner reduction. However, as shown in next section, the paradox may also occur with belief functions, depending on the reduction, and without being triggered necessarily in the same time as with a Bayesian approach.

5 Numerical Simulations

In this section, the results of the following experiment are given:

- 10^8 contingency tables of 3 binary variables composed of numbers defined randomly in the interval $[1, 10^3]$ are built. Frequencies about the paradox are then observed for the three approaches: probability case, belief functions with inner and outer reduction cases.
- the preceding point is repeated 10 times in order to obtain mean values and variances of the obtained frequencies.

Table 4 provides the mean and standard deviation values obtained for the eight frequencies of the paradox triggering in the experiment. Each frequency corresponds to a case of appearance considering the probability approach (*Proba*) and the approaches with belief functions marginalized with the inner and outer reductions (*Inner*, *Outer*). A "1" means a paradox.

A first remark when looking at these results is that Simpson's paradox can be triggered for the three approaches, and all combinations of occurrences are possible.

Table 4. Appearance frequencies of Simpson's paradox over 10^8 contingency tables composed of numbers randomly chosen between 1 and 1000. "1" means a paradox. The first number is the mean value, the second one the standard deviation.

	$Proba = 0$		$Proba = 1$	
	$Outer = 0$	$Outer = 1$	$Outer = 0$	$Outer = 1$
$Inner = 0$	94% 0.003%	3.77% 0.0012%	0.04% $1.5.10^{-4}$%	0.9% 7.10^{-4}%
$Inner = 1$	1.2% 0.0014%	0.004% 9.10^{-5}%	0.013% $1.4.10^{-4}$%	0.0046% 9.10^{-5}%

In most cases (94%), the paradox is absent for all the approaches.

Simpson's paradox occurs in this experiment for the probability approach with a 0.9576% frequency[1] $(0.04 + 0.9 + 0.013 + 0.0046)$ which is lower than for the other cases (inner and outer reductions).

At last it can be observed that the chance (0.9%) of observing a paradox for both the probabilistic and outer reduction approaches, is much more important than that of observing a paradox for the inner reduction approach with any or both of the other approaches. Taking into account the marginal chances of observing a paradox for each approach, we may further remark that the probabilistic approach is actually somewhat included in the outer reduction approach (when a paradox is observed for the probabilistic approach, a paradox will in general be observed for the outer reduction), whereas the inner reduction approach is almost disjoint from the other ones.

6 Conclusion and Discussion

In this article, similarly to the probability case, frequencies of appearance of Simpson's paradox have been studied with belief functions. The marginalization step is shown to impact it, since the inner and outer reductions have different behaviours. It has been shown than a paradox can occur in each approach and each combination of approaches.

To complete this study, instead of the least specific isopignistic belief function, it might be interesting to investigate on other estimators since they may yield different results than those obtained with consonant beliefs. Indeed, for instance, the Bayesian belief function identified to the probability measure, leads to different conclusions than those based on consonant beliefs.

Another parameter that influences the paradox is the decision rule. In this paper the pignistic rule is used. But other decision rules exist, and among them

[1] The same frequency was valued at 1.67% in another experiment [8]. In this article, the experiment is based on 7 independent proportions, the eight one summing up to one. This discrepancy between the two results may be due to the fact that proportions may shift the triggering of the paradox, as recalled in the penultimate paragraph of Section 2.

the maximum of plausibility emerges. However, the analysis of the full decision chain based on this rule is more complex (for instance, differences between decisions obtained under this rule and under the pignistic rule already appear on the set Ω) and is left for further work.

Lastly, it would be interesting to investigate whether causality as addressed in belief function theory [1], could bring a solution to the paradox, as is in the probabilistic case.

References

1. Boukhris, I., Elouedi, Z., Benferhat, S.: Dealing with external actions in belief causal networks. International Journal of Approximate Reasoning 54(8), 978–999 (2013)
2. Chater, N., Vlaev, I., Grinberg, M.: A new consequence of Simpson's Paradox: Stable co-operation in one-shot Prisoner's Dilemma from populations of individualistic learning agents, University College London/New Bulgarian University, `http://else.econ.ucl.ac.uk/papers/uploaded/210.pdf`
3. Cobb, B.R., Shenoy, P.P.: On the plausibility transformation method for translating belief function models to probability models. International Journal of Approximate Reasoning 41(3), 314–330 (2006)
4. Dubois, D., Prade, H.: A set-theoretic view of belief functions: logical operations and approximations by fuzzy sets. International Journal of General Systems 12(3), 193–226 (1986)
5. Dubois, D., Moral, S., Prade, H.: Belief Change Rules in Ordinal and Numerical Uncertainty Theories. In: Handbook of Defeasible Reasoning and Uncertainty Management Systems, vol. 3, pp. 311–392 (1998)
6. Dubois, D., Prade, H., Smets, P.: A Definition of Subjective Possibility. International Journal of Approximate Reasoning 48, 352–364 (2008)
7. Malinas, G.: Simpson's Paradox, Stanford encyclopedia of philosophy, `http://plato.stanford.edu/entries/paradox-simpson/` (first published February 2, 2004)
8. Pavlides, M.G., Perlman, M.D.: How likely is Simpson's paradox? The American Statistician 63(3), 226–233 (2009)
9. Pearl, J.: Causality: Models, Reasoning, and Inference. Cambridge University, New York (2000)
10. Sentz, K., Ferson, S.: Combination of evidence in Dempster–Shafer theory, Technical report, SANDIA National Laboratories, 96 pages (2002)
11. Simpson, E.H.: The interpretation of interaction in contingency tables. Journal of the Royal Statistical Society, Ser. B 13, 238–241 (1951)
12. Simpson's paradox (2008), `http://en.wikipedia.org/wiki/Simpson's_paradox`
13. Shafer, G.: A mathematical theory of evidence. Princeton University Press, Princeton (1976)
14. Smets, P., Kennes, R.: The transferable belief model. Artificial Intelligence 66, 191–234 (1994)
15. Smets, P.: Decision Making in the TBM: the Necessity of the Pignistic Transformation. International Journal of Approximate Reasoning 38(2), 133–147 (2005)
16. Approaches to Causal Structure Learning, Workshop, UAI 2012-2013, `http://www.stat.washington.edu/tsr/uai-causal-structure-learning-workshop/`, and all papers available online