Inferring Causal Directions in Errors-in-Variables Models *

Yulai Zhang and **Guiming Luo**
School of Software, Tsinghua University
Beijing, 100084, China P.R.
{zhangyl08@mails, gluo@mail}.tsinghua.edu.cn

Abstract

A method for inferring causal directions based on errors-in-variables models where both the cause variable and the effect variable are observed with measurement errors is concerned in this paper. The inference technique and estimation algorithms are given. Some experiments are included to illustrate our method.

Introduction

Causal discovery is an essential topic in artificial intelligence. Inferring the correct causal direction between two correlated variables is the most basic but nontrivial problem in this field. Generally speaking, causal directions can be decided by the properly designed experiments. However, when such controlled experiments are impossible to be performed, it is necessary to discover the causal relationships from the observed data sets.

Various models are proposed to infer the causal directions between two observed data sets in recent years. Most of these models assume that at least one of the variables in a causal-effect pair is accurately measured (Janzing et al. 2012; Hoyer et al. 2008). However, observations with measurement errors for both the input and output of a natural system are common and make the problems more difficult.

In several other models (Shimizu et al. 2006; Zhang and Luo 2013), though all the variables are observed with errors, the noises contaminate the input cause variables and affect the models' output. Thus, asymmetries are introduced in these models by this assumption, and the cause effect directions can be inferred by these asymmetries.

In our point of view, many natural processes are more like the errors-in-variables models where both the causal and effect variables are observed with the measurement noises and the measurement noises are not input to the natural systems, just as show in (1)-(3). Thus, investigating the causal relationship in the EIV models is meaningful. In addition, since the input noise does not affect the output, this symmetric nature of the EIV model also makes it a challenging task.

*This work was supported by the Funds NSFC61171121 and the Science Foundation of Chinese Ministry of Education - China Mobile 2012.

An EIV model can be described as:

$$y(i) = f(x(i)) \tag{1}$$

$$y_o(i) = y(i) + e_y(i) \tag{2}$$

$$x_o(i) = x(i) + e_x(i) \tag{3}$$

where $\{x_o(i), y_o(i)\}$ $i = 1...N$ is the observed data set.

In this work, we will focus on the causal direction inference in EIV models.

Inference Analysis

Many natural systems can be described as (1). The output variable y bares the information of the input variable x and the information of the transfer function f. If the probability density function of x is $p_x(x)$, the pdf of y can be written as:

$$p_y(y) = p_x(g(y))|g'(y)| \tag{4}$$

where function g is the inverse function of f, $x(i) = g(y(i)) = f^{-1}(y(i))$. In some of the following discussions, the index variable i will be omitted for simplicity.

When the distribution of the cause variable x and the transfer function f are chosen independently, the distribution of the effect variable y and the derived function of g will be correlated. This can be intuitively observed from (4). Thus, the causal direction can be decided by comparing the degrees of correlations.

Note that if f is linear, the asymmetry between the two directions will disappear because both $f'(x)$ and $g'(y)$ are constant. In this work, we restrict our discussion on nonlinear or piecewise linear transfer functions. In addition, we assume that e_x and e_y have zero means and normal distributions, and are independent with the corresponding actual variables. Measurement noises which are consistent with these assumptions occur quite often in practice.

In the following subsections we will investigate: 1)how to calculate the correlation between functions; 2)how to obtain the transfer functions, and 3)how to estimate the pdf of the actual variables under measurement noises.

Let $f_1(t)$ and $f_2(t)$ be defined and integrable in the observation interval $[a, b]$. The interval can be determined by the maximum and minimum value of the observations. The correlation coefficient ρ can be calculated as:

$$\rho(f_1, f_2) = \langle \bar{f}_1(t), \bar{f}_2(t) \rangle \tag{5}$$

where $\langle\cdot,\cdot\rangle$ is the inner product operator and the values of the functions are normalized in the interval $[a,b]$ by $\bar{f}(t) = (f(t) - mean(f(t)))/\|f(t) - mean(f(t))\|_2$.

In order to infer the causal direction, we chose $f_1(t) = |f'(t)|$, $f_2(t) = p_x(t)$ for ρ_{in}, and $f_1(t) = |g'(t)|$, $f_2(t) = p_y(t)$ for ρ_{out}. The causal directions can be determined by the fact that $|\rho_{in}| < |\rho_{out}|$.

Estimate Algorithms

Regression is a core research topic in many research fields. The regression for EIV models encounters extra difficulties. Since the derived functions of f and g are needed, a parametric model with explicit formulations are more suitable for our task. We expand the function f in basis functions φ_d, and let $f(x) = \sum_{d=0}^{n} \theta_d \varphi_d(x)$. For EIV model the parameter vector θ can be obtained by:

$$\min_{\theta} J_1 = \theta^T(E\varphi(x_o)\varphi(x_o)^T)\theta(\theta^T C\theta)^{-1} \tag{6}$$

where $\varphi(x_o) = [y_o, \varphi_0(x_o), ..., \varphi_d(x_o)]^T$, and $\theta = [-1, \theta_0, ..., \theta_d]^T$. An explicit estimator for the polynomial basis was proposed in (Vajk and Hetthéssy 2003). This algorithm requires a known covariance matrix C which can be determined by $\sigma_{e_x}^2$ and $\sigma_{e_y}^2$ in our model. From (2) and (3), we can obtain the probability density function by: $p_{x_o}(x) = \int p_x(x-\varepsilon)p_\varepsilon(\varepsilon)d\varepsilon$. The pdf of x_o is the convolution of pdf of x and e_x. The deconvolution can be calculated by the pdfs' characteristic functions:

$$\phi_x(t) = \phi_{x_o}(t)/\phi_{e_x}(t) \tag{7}$$

where the characteristic function $\phi_x(t) = Ee^{itx}$ can be calculated by FFT. Since we have assumed a zero mean gaussian density, the rest problem is how to obtain the variances of the noises e_x and e_y.

A noise variance estimate criterion based on instrumental variable for linear dynamic EIV models is proposed in (Diversi, Guidorzi, and Soverini 2006). This criterion can be extended to polynomial nonlinear models as:

$$\min_{\sigma_{e_x}^2} J_2 = \theta^T(E(\tilde{\varphi}\bar{\varphi}^T)^T)E(\tilde{\varphi}\bar{\varphi}^T)\theta \tag{8}$$

$$s.t.\sigma_{e_y}^2 = R_{yy} - R_{yx}(R_{xx} - E\varphi(e_x)\varphi(e_x)^T)^{-1}R_{yx}^T$$

where $R_{yx} = Ey_o[\varphi_1(x_o), ..., \varphi_d(x_o)]$, $R_{yy} = Ey_o y_o^T$ and $R_{xx} = E[\varphi_1(x_o), ..., \varphi_d(x_o)]^T[\varphi_1(x_o), ..., \varphi_d(x_o)]$ can be calculated directly. $\bar{\varphi} = [y, \varphi_0(x), ..., \varphi_d(x)]^T$, and $\tilde{\varphi} = [\varphi_{d+1}(x), ..., \varphi_{d+q}(x)]^T$ is the instrumental variable. Note that although the accurate values of x, y and e_x are not available, $E(\varphi(e_x)\varphi(e_x)^T)$ and $E\tilde{\varphi}\bar{\varphi}^T$ can be presented by the observation x_o, y_o and $\sigma_{e_x}^2, \sigma_{e_y}^2$, so $\sigma_{e_x}^2$ is the only independent variable in (8).

Algorithm and Experiments

We propose an EM style algorithm in this work to solve (6) and (8) together as follow:

Step 1. Select the model order n and the length of the instrumental variable q. Initialize the values of the optimization variables in (6) and (8).

Step 2. Solve (6), get the result of $\theta(t)$ at tth step.

Step 3. Solve (8), and get $\sigma_{e_x}^2(t)$. While $|\sigma_{e_x}^2(t)/\sigma_{e_x}^2(t-1) - 1| > \delta$, go back to Step 2.

Step 4. Calculate the pdf $p_x(x)$ using (7)

Step 5. Calculate ρ_{xy} by (5)

Step 6. Exchange x and y, do Step 2 to Step 5 again to calculate ρ_{yx}.

Step 7. If $|\rho_{xy}| < |\rho_{yx}|$, the causal direction is x causes y, and if $|\rho_{yx}| < |\rho_{xy}|$, the causal direction is y causes x.

In this experiment, we compare the method in this paper with the entropy based IGCI method proposed in (Janzing et al. 2012). The data is from second order polynomial models. θ_d is randomly generated between [0,2]. The input data x is generated by a uniform distributed between -0.5 and 1.5. $\sigma_{e_x}^2$ and $\sigma_{e_y}^2$ are equally selected from 0.1 to 0.3. The percentage of successful causal direction inference for both methods are shown in the fig.1. The method proposed in this paper gives better results for data with measurement errors.

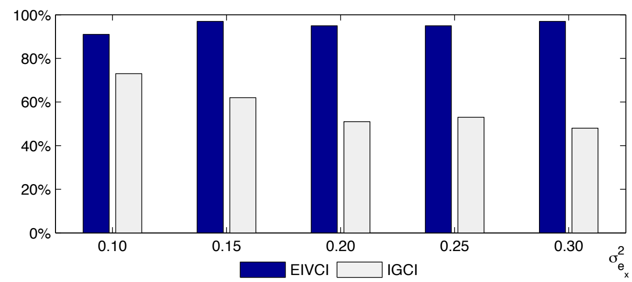

Figure 1: Experiment Result of EIVCI (EIV Causal Inference) and the IGCI method.

References

Diversi, R.; Guidorzi, R.; and Soverini, U. 2006. Yule-walker equations in the frisch scheme solution of errors-in-variables identification problems. In *Proc. 17th international symposium on mathematical theory of networks and systems. MTNS. Kyoto, Japan. July*, 391–395.

Hoyer, P. O.; Janzing, D.; Mooij, J.; Peters, J.; and Scholkopf, B. 2008. Nonlinear causal discovery with additive noise models. In *Proceedings of The Twenty-Second Annual Conference on Neural Information Processing Systems (NIPS 2008)*.

Janzing, D.; Mooij, J.; Zhang, K.; Lemeire, J.; Zscheischler, J.; Daniušis, P.; Steudel, B.; and Schölkopf, B. 2012. Information-geometric approach to inferring causal directions. *Artificial Intelligence* 182:1–31.

Shimizu, S.; Hoyer, P. O.; Hyvarinen, A.; and Kerminen, A. 2006. A linear non-gaussian acyclic model for causal discovery. *Journal of machine learning reseach* 7:2003–2030.

Vajk, I., and Hetthéssy, J. 2003. Identification of nonlinear errors-in-variables models. *Automatica* 39(12):2099–2107.

Zhang, Y., and Luo, G. 2013. An entropy based method for causal discovery in linear acyclic model. In *Neural Information Processing, LNCS*, 250–256. Springer.