NeuroImage xxx (2015) xxx–xxx

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg

Identification of causal relations in neuroimaging data with latent confounders: An instrumental variable approach☆

Moritz Grosse-Wentrup [a,*], Dominik Janzing [a], Markus Siegel [b], Bernhard Schölkopf [a]

[a] Empirical Inference Department, Max Planck Institute for Intelligent Systems, Spemannstr. 38, 72076 Tübingen, Germany
[b] Centre for Integrative Neuroscience and MEG Center, Eberhard-Karls Universität Tübingen, Otfried-Müller-Str. 47, 72076 Tübingen, Germany

A R T I C L E I N F O

Article history:
Received 24 April 2015
Accepted 13 October 2015
Available online xxxx

A B S T R A C T

We consider the task of inferring causal relations in brain imaging data with latent confounders. Using a priori knowledge that randomized experimental conditions cannot be effects of brain activity, we derive statistical conditions that are sufficient for establishing a causal relation between two neural processes, even in the presence of latent confounders. We provide an algorithm to test these conditions on empirical data, and illustrate its performance on simulated as well as on experimentally recorded EEG data.

Introduction

Inferring the causal structure of a cortical network is a central goal in neuroimaging (Smith et al., 2011). Various methods have been developed to infer causal relations from brain imaging data, including structural equation modeling (SEM) (Mcintosh and Gonzalez-Lima, 1994; Atlas et al., 2010), Granger causality (GC) (Granger, 1969; Kamiński et al., 2001; Gregoriou et al., 2009), dynamic causal modeling (DCM) (Friston et al., 2003; Daunizeau et al., 2011), and causal Bayesian networks (CBNs) (Ramsey et al., 2010; Grosse-Wentrup et al., 2011; Ramsey et al., 2011; Mumford and Ramsey, 2014; Weichwald et al., 2015). These methods commonly assume *causal sufficiency*; that is, they presume that all causally relevant variables have been observed. This assumption is often implausible, because various factors can confound a causal analysis. These factors include, but are not limited to, unmeasured brain regions in an fMRI analysis (Mcintosh and Gonzalez-Lima, 1994; Daunizeau et al., 2011; Friston et al., 2011), cardio-ballistic artifacts in ECoG recordings (Kern et al., 2013), and volume conduction of cortical and non-cortical current sources in EEG or MEG data (Grosse-Wentrup, 2009; Hipp and Siegel, 2013). Because it is not trivial to anticipate potential confounders, results obtained with methods based on causal sufficiency must be interpreted with caution.

Latent confounders can be addressed by the IC* (Pearl, 2000) and FCI algorithms (Spirtes et al., 2000; Zhang, 2008), which use the theory of ancestral graphs. Theoretically, both algorithms can distinguish genuine causal relations from spurious relations induced by latent confounders. In practice, the involved statistical tests are complex, which currently limits their application in neuroimaging to variables that are jointly Gaussian distributed (Waldorp et al., 2011). The assumption of jointly Gaussian distributed variables has been criticized as unreasonable for neuroimaging data (Hanson and Bly, 2001; Wink and Roerdink, 2006; Mumford and Ramsey, 2014).

We contribute to research on causal inference with latent confounders in two ways. First, we show that the statistical tests required to identify a genuine causal relation can be simplified when the experimental condition is randomized. Using the a priori knowledge that a randomized experimental condition cannot be caused by neural processes, we analytically prove that if two neural processes are modulated by an experimental condition, a single test of conditional independence is sufficient to establish a genuine causal relation between those processes. To emphasize the requirement that, in our approach, the experimental conditions must be randomized, we later refer to them as the stimuli presented to a subject. Second, by using linear regression, we reduce the required conditional independence test to a marginal independence test. This test is advantageous because asymptotically consistent statistical tests are readily available for marginal independence (Gretton et al., 2005, 2008; Gretton and Györfi, 2010), but not for conditional independence (Fukumizu et al., 2008; Zhang et al., 2011). We prove that this linearized conditional independence test is sufficient but not necessary for conditional independence: while our test may fail to detect conditional independence if the assumption of linearity is not met, a positive test result implies that this assumption has been fulfilled. Taken together, our two contributions lead to a non-parametric version of the instrumental variable approach to causal inference (Angrist et al., 1996; Pearl, 2000). The resulting algorithm, which we term *stimulus-based causal inference* (SCI), can provide empirical evidence for a causal relation between two neural processes, even in the presence of latent confounders.

☆ M.S. was supported by the Centre for Integrative Neuroscience (Deutsche Forschungsgemeinschaft, EXC 307). The authors would like to thank Sebastian Weichwald for comments on an earlier version of this article.

* Corresponding author.
E-mail address: moritzgw@tuebingen.mpg.de (M. Grosse-Wentrup).

http://dx.doi.org/10.1016/j.neuroimage.2015.10.062
1053-8119/

We demonstrate the performance of the SCI algorithm on simulated as well as on experimentally recorded EEG data. We first use a neural mass model for spectral responses in electrophysiology (Moran et al., 2007) to provide estimates of the power and of the false discovery rate (FDR) of the SCI algorithm for a variety of causal models. We then show how our method can be used to infer group-level causal relations on EEG data, which we recorded for a study on brain–computer interfacing (BCI) (Grosse-Wentrup and Schölkopf, 2014). In this study, subjects were trained via neurofeedback to self-regulate the amplitude of γ-oscillations (55–85 Hz) in the right superior parietal cortex (SPC), a primary node of the central executive network (CEN) (Bressler and Menon, 2010). Because transcranial magnetic stimulation (TMS) of the CEN has been found to modulate the medial prefrontal cortex (MPC) (Chen et al., 2013), we hypothesized that self-regulation of γ-power in the right SPC causes variations in γ-power in the MPC. Consistent with this hypothesis, the SCI algorithm determined the MPC to be modulated by the right SPC. We conclude the article with a discussion of the utility and of the limitations of causal inference to study the structure and the function of cortical networks.

We note that the SCI algorithm is applicable not only to EEG recordings but also to any neuroimaging data set that is based on randomized experimental conditions. We have condensed the SCI algorithm into one line of Matlab code, which is available at http://brain-computer-interfaces.net.

Methods

We begin this section by introducing the framework of causal Bayesian networks (CBNs), which our work is based on (cf. Ramsey et al., 2010; Grosse-Wentrup et al., 2011; Ramsey et al., 2011; Mumford and Ramsey, 2014; Weichwald et al., 2015 for applications of this framework in neuroimaging). We then present the sufficient conditions to establish causal influence of one cortical process on another in stimulus-based experiments (Section 2.2). In Section 2.3, we use linear regression to reduce the required conditional independence test to a marginal independence test. We discuss how to apply the resulting causal inference procedure to empirical data in Section 2.4. We conclude the methods section with a discussion of the relation of the SCI algorithm to instrumental variables in Section 2.5.

Causal Bayesian networks

In the framework of CBNs, a random variable x is a cause of another random variable y if setting x to different values by an external intervention changes the probability distribution over y (Pearl, 2000; Spirtes et al., 2000). In the notation of the *do*-calculus, this is expressed as $p(y|\mathrm{do}(x)) \neq p(y)$ for some values of x and y. Thus, the framework of CBNs defines cause-effect relations in terms of the impact of external manipulations. This definition contrasts those of frameworks which define causality in terms of information transfer (Granger, 1969; Roebroeck et al., 2005; Gregoriou et al., 2009; Lizier and Prokopenko, 2010).

Causal relations between a set $\mathcal{X}$ of random variables are represented by edges in a directed acyclic graph (DAG). The *causal Markov condition* (CMC) relates the structure of a DAG, as represented by its edges, to statistical independence relations between the variables in $\mathcal{X}$. Specifically, it states that every (conditional) independence implied by a DAG is also found in the joint probability distribution $p(\boldsymbol{x})$. We recall that two random variables x and y are statistically independent (conditional on a third random variable z) if and only if their joint distribution factorizes into the product of its marginals, i.e. if and only if $p(x, y) = p(x)p(y) \quad (p(x, y|z) = p(x|z)p(y|z))$. Intuitively, this states that observing x does not provide any information on how likely certain outcomes of y are (and vice versa). We abbreviate statistical independence between x and y (conditional on z) as $x \perp\!\!\!\perp y (x \perp\!\!\!\perp y|z)$. Assuming the CMC, (conditional) independence relations can be read off the structure of a DAG by checking for *d-separation* properties. A set of nodes $\mathcal{D}$ is said to d-separate x and y if every path from x to y contains at least one variable z such that either z is a collider ($\rightarrow z \leftarrow$) and no descendant of z (including z itself) is in $\mathcal{D}$; or z is not a collider and z is in $\mathcal{D}$. We provide examples of d-separation in the next paragraph and refer the interested reader to Pearl (2000) or Spirtes et al. (2000) for a more exhaustive introduction to the concept of d-separation. The CMC thus relates structural properties of DAGs to empirically observable independence relations. To perform causal inference, we also need to relate empirically observable independence relations to structural properties of the data-generating DAG. This is achieved by the assumption of *faithfulness*. Faithfulness asserts that every (conditional) independence relation in $p(\boldsymbol{x})$ is implied by the structure of the associated DAG. Taken together, the CMC and faithfulness ensure that two variables x and y are conditionally independent given z if and only if x and y are d-separated by z. This equivalence gives us insight into the structure of a DAG from empirically testable (conditional) independence relations.

We now provide three examples of d-separation that are relevant to our following arguments. First, consider the chain $x \rightarrow z \rightarrow y$. Here, x and y are marginally dependent ($x \not\perp\!\!\!\perp y$), because x influences y via z. However, as z d-separates x and y by blocking the directed path from x to y, x and y are statistically independent given $z (x \perp\!\!\!\perp y|z)$. Second, consider the fork $x \leftarrow z \rightarrow y$. Again, x and y are marginally dependent ($x \not\perp\!\!\!\perp y$), because they share a common cause z. This common cause z again d-separates x and y by removing the joint effect of z on x and y, rendering x and y independent conditional on $z (x \perp\!\!\!\perp y|z)$. Third, consider the collider $x \rightarrow z \leftarrow y$. In this case, x and y are independent ($x \perp\!\!\!\perp y$), because they are d-separated by the empty set. Because z is a joint effect of x and y, however, it unblocks the previously blocked path between x and y, rendering x and y dependent conditional on ($x \not\perp\!\!\!\perp y|z$).

These three examples form the basis of causal inference in CBNs. For instance, if we observe that $x \perp\!\!\!\perp y$ yet $x \not\perp\!\!\!\perp y|z$, then we can conclude that our data has not been generated by a chain or by a fork. These observations limit the possible causal structures to only collider and DAGs with additional (latent) variables. A more comprehensive introduction to the framework of CBNs in the context of neuroimaging is given in Mumford and Ramsey (2014).

Causal inference in stimulus-based paradigms

In this article, we only consider DAGs over a set of three random variables, $\mathcal{V} = \{s, x, y\}$. The variables x and y represent brain state features, and s represents an experimental condition. For our theoretical arguments, we assume the joint probability distribution $p(s, x, y)$ to be known. This assumption implies that we have access to an oracle for any conditional independence relation in $\mathcal{V}$. We relax this assumption in Section 4. Note that, while x and y may represent any measure of brain activity, it is helpful to consider trial-averaged blood-oxygen-level-dependent (BOLD) activity at different cortical locations or trial-averaged band power at two EEG channels as examples.

In the following, we assume that s codes a randomized experimental stimulus that is presented to the subject before x and y are measured. This assumption leads to the following theorem.

Theorem 1. Causal inference in stimulus-based paradigms

Let s, x, and y be three random variables with a joint probability distribution $p(s, x, y)$ that is faithful to its generating DAG. Further, assume that s codes a randomized experimental stimulus that is presented before x and y are measured. Then the following three conditions are sufficient for x to be a genuine cause of y ($x \rightarrow y$):

1. *s is not independent of ($s \not\perp\!\!\!\perp x$),*
2. *s is not independent of y ($s \not\perp\!\!\!\perp y$), and*
3. *s and y are independent conditional on $x (s \perp\!\!\!\perp y|x)$.*