

MIT Open Access Articles

A general statistical framework for assessing Granger causality

The MIT Faculty has made this article openly available. **Please share** how this access benefits you. Your story matters.

Citation: Kim, Sanggyun, and Emery N. Brown. "A General Statistical Framework for Assessing Granger Causality." in Proceedings of the 2010 IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP), IEEE, Sheraton Dallas Hotel, Dallas, TX, 14-19 March 2010. 2222-2225. Web. ©2010 IEEE.

As Published: <http://dx.doi.org/10.1109/ICASSP.2010.5495775>

Publisher: Institute of Electrical and Electronics Engineers

Persistent URL: <http://hdl.handle.net/1721.1/70484>

Version: Final published version: final published article, as it appeared in a journal, conference proceedings, or other formally published context

Terms of Use: Article is made available in accordance with the publisher's policy and may be subject to US copyright law. Please refer to the publisher's site for terms of use.



A GENERAL STATISTICAL FRAMEWORK FOR ASSESSING GRANGER CAUSALITY

Sanggyun Kim¹ and Emery N. Brown^{1,2,3}

¹Department of BCS, ²Harvard-MIT Division of HST, MIT, Cambridge, MA 02139, USA

³Neuroscience Statistics Research Lab, MGH, Harvard Medical School, Boston, MA 02114, USA
{sgkim, enb}@neurostat.mit.edu

ABSTRACT

Assessing the causal relationship among multivariate time series is a crucial problem in many fields. Granger causality has been widely used to identify the causal interactions between continuous-valued time series based on multivariate autoregressive models in the Gaussian case. In order to extend the application of the Granger causality concept to non-Gaussian time series, we propose a general statistical framework for assessing the causal interactions. In this study, the Granger causality from a time series x_2 to a time series x_1 is assessed based on the relative reduction of the likelihood of x_1 by the exclusion of x_2 compared to the likelihood obtained using all the time series. Simulation results indicated that the proposed algorithm accurately predicted nature of interactions between discrete-valued time series as well as between continuous-valued time series.

Index Terms— Granger causality, generalized linear model, false discovery rate, neural spike train data

1. INTRODUCTION

Identifying causal interactions between time series is a crucial problem in many fields such as economics, social science, and physics. Recently in neuroscience the ability to exploit causal relationships, as opposed to just correlation, is also becoming more important to the understanding of cooperative nature of neural computation [1], [2]. The basic idea of the causality between two time series was introduced by Wiener [3] but was too general to implement practically. Granger formalized this idea in order to enable practical implementation based on the multivariate autoregressive (MVAR) models [4]: if the variance of the prediction error of a time series $x_1(t)$ is reduced by the inclusion of the past values of a time series $x_2(t)$, then x_2 is said to cause x_1 in a ‘Granger’ sense. In the case of more than two time series, causal relationship between any two of the series may be indirect, i.e., mediated by another time series, and this issue was addressed by the technique called conditional Granger causality [5].

This work was supported by NIH Grants DP1-OD003646 and R01-EB006385.

Linear Granger causality (LGC) based on the MVAR models, combined with an implicit Gaussianity assumption on the time series, provides a widely used framework for testing causality between continuous-valued time series [5], [6]. However, it is applicable only to Gaussian time series, but inapplicable to non-Gaussian case such as discrete-valued time series.

To address this issue, we propose a general statistical framework for assessing the causal relationships. When the probability of a time series at the current time is modeled using the past values of itself and other time series, we assess the Granger causality based on the likelihood ratio test [6]. That is, we assess the Granger causality from a time series x_2 to x_1 by calculating the relative reduction of the likelihood of x_1 by the exclusion of x_2 compared with the likelihood obtained using all the time series. If the likelihood ratio is less than one, there is causal influence from x_2 to x_1 , and if the ratio is one, there is no causality. The Granger causality measure based on the likelihood ratio test is exactly reduced to the LGC measure in the Gaussian case [6]. The likelihood ratio test statistic also enables us to perform several statistical significance tests such as the chi-squared test, the Z-test, and the F-test, which investigate the original causal connectivity between multiple time series. When assessing the causality relationships between multiple time series, we use the false discovery rate (FDR) to correct for multiple hypothesis testing [8]. We also show how the dependence measure between two time series x_1 and x_2 can be decomposed based on the proposed measure.

Simulation results demonstrate that the proposed framework estimated the underlying causal connectivity more exactly rather than the LGC for discrete-valued and mixture model case.

2. PROBABILISTIC GRANGER CAUSALITY

Consider K multivariate time series $\{x_k(t)\}_{k=1}^K$ for $0 \leq t \leq T$. Let us denote the past values of a time series $x_k(t)$ at time t as $\underline{x}_k(t) = \{x_k(s), s < t\}$, the past values of all time series at time t as a column vector $\underline{\mathbf{x}}(t) = [\underline{x}_1(t), \dots, \underline{x}_K(t)]^T$, and the past values obtained after excluding $\underline{x}_j(t)$ from $\underline{\mathbf{x}}(t)$

as $\underline{x}^j(t)$, respectively. Let us assume that x_i is *not* caused by x_j , i.e., x_j has no direct influence on x_i or indirect influence that is mediated by other time series. In this case, $x_i(t)$ is conditionally independent of $\underline{x}_j(t)$ given all the past values that have direct influences on $x_i(t)$ [7]. This can be mathematically represented as

$$p(x_i(t)|\underline{x}(t)) = p(x_i(t)|\underline{x}^j(t)). \quad (1)$$

When the left-hand side of the equation (1) is greater than the right-hand side, x_j is said to cause x_i .

Based on this property, we assess the Granger causality using the log-likelihood ratio. From (1), the likelihood function of time series x_i is represented as

$$L_i(\theta_i|x_i) = \int_0^T p(x_i(t)|\underline{x}(t); \theta_i) dt \quad (2)$$

where the parameter vector θ_i includes information about the dependency of $x_i(t)$ on $\underline{x}(t)$. Henceforth, we will denote $L_i(\theta_i|x_i)$ as $L_i(\theta_i)$ for simplicity. Using the likelihood of (2), the Granger causality from x_j to x_i is assessed as

$$\Gamma_{ij} = \log \frac{L_i(\theta_i^j)}{L_i(\theta_i)} \quad (3)$$

where the likelihood $L_i(\theta_i^j)$ is calculated using $\underline{x}^j(t)$ instead of $\underline{x}(t)$ of (2) in order to remove the effect of x_j on x_i and the parameter vector θ_i^j for $\underline{x}^j(t)$. If x_j causes x_i in the Granger sense, the likelihood $L_i(\theta_i)$ is greater than the likelihood $L_i(\theta_i^j)$, i.e., $\Gamma_{ij} < 0$. In contrast, if x_j does not cause x_i , the likelihood ratio of (3) is equal to one, i.e., $\Gamma_{ij} = 0$.

3. SIGNIFICANCE TEST

The Granger causality measure of (3) represents the relative strength of causal interactions among time series but provides little insight into the original connectivity among them. To investigate the original causal connectivity, our study introduces statistical significance test based on the log-likelihood ratio statistic.

The significance test is based on the goodness-of-fit (GOF) statistics [9]. Consider the null hypothesis given as $H_0 : \theta_0 = \theta_i^j$. An alternative hypothesis is given as $H_1 : \theta_1 = \theta_i$. These two models are nested since a model for θ_i^j is a special case of the more general model for θ_i . We can test H_0 against H_1 using the test statistic S , which is given by $-2\Gamma_{ij}$. If these models describe the data well, then the test statistic S will be asymptotically chi-squared, which is described simply as $S \sim \chi_M^2$ where the degree of freedom M is the difference in dimensionality of two models [9].

If the value of the test statistic S is consistent with χ_M^2 distribution, H_0 is accepted since it is simpler. This result

indicates that $\underline{x}_j(t)$ contains no significant information that would assist in predicting $x_i(t)$. Thus, x_j has no causal influence on x_i . On the contrary, if the value of S is in the critical region, i.e., greater than the upper tail $100 \times (1 - \alpha)\%$ of the distribution, then H_0 may be rejected in favor of H_1 since the model of H_1 describes the data with significantly more accuracy. This indicates that $\underline{x}_j(t)$ contains information that improves the ability to predict $x_i(t)$. Thus x_j causes x_i in the Granger sense. In this work, α was set to 0.05.

When the generalized linear model (GLM) framework is used to model the distribution of times series, we can use the deviance as the GOF statistic [10]. The deviance difference between two models is given by $\Delta D = -2\Gamma_{ij}$, which is used as the GOF statistic. However, for the Gaussian distribution the deviance depends on the variance of the distribution, which is usually unknown in practice. In this case in order to remove the variance term the ratio $S = \Delta D/D_1$ where D_1 is the deviance obtained based on θ_1 is used as the test statistic, which can be now calculated from the data. If H_0 is correct, S follows the central F-distribution, $F(M, N - p)$ where N is the number of observations and p is the dimensionality of θ_1 .

When assessing the causal interactions between multiple time series, we should consider a family of hypothesis inferences simultaneously; however, multiple hypothesis tests are more likely to incorrectly reject the null hypothesis [11]. In our study, we control the expected proportion of incorrectly rejected null hypotheses (type I errors) by exploiting the FDR [8].

4. DECOMPOSITION OF DEPENDENCE MEASURE

In our study, we show how the dependence measure between two time series can be decomposed based on the Granger causality measure based on the likelihood ratio statistic. Symmetrically to (3), the Granger causality measure from x_i to x_j is obtained as

$$\Gamma_{ji} = \log \frac{L_j(\theta_j^i)}{L_j(\theta_j)}. \quad (4)$$

In addition to two unidirectional Granger causality measures, we can consider another causality measure called instantaneous causality, which may be caused by an exogenous common input or a low temporal resolution of data. The instantaneous causality measure between x_i and x_j is defined as

$$\Gamma_{i,j} = \log \frac{L_i(\theta_i)L_j(\theta_j)}{L_{ij}(\theta_{ij})} \quad (5)$$

where $L_{ij}(\theta_{ij})$ is the joint likelihood function of x_i and x_j . If $\Gamma_{i,j} \neq 0$, then x_i and x_j instantaneously cause one another.

If x_i and x_j are independent of one another, the joint density function can be decomposed into the product of marginal

density functions of x_i and x_j . So we can define the measure of dependence between x_i and x_j as

$$\Gamma_{i,j} = \log \frac{L_i(\theta_i^j)L_j(\theta_j^i)}{L_{ij}(\theta_{ij})}. \quad (6)$$

Using the definitions cited above, we obtained

$$\Gamma_{i,j} = \Gamma_{ij} + \Gamma_{ji} + \Gamma_{i,j}. \quad (7)$$

Therefore, the total dependence between two time series x_i and x_j is decomposed into three measures: two unidirectional Granger causality measures between x_i and x_j and one non-directional instantaneous causality measure between them.

5. SIMULATIONS

In this paper, we used the GLM framework to model the distribution of time series [10]. In a GLM framework, each time series is assumed to be generated from a particular distribution function in the exponential family, which allows us to model count data by Poisson, binomial, and multinomial distributions as well as skewed data by Gamma distributions. We analyzed the Granger causality in the GLM framework based on an assumption that the mean of each time series depends on its own history and the past values of concurrent time series as

$$E[x_i(t)] = \mu_i(t|\mathbf{x}(t)) = g_i^{-1}(\mathbf{x}(t)^T \boldsymbol{\theta}_i) \quad (8)$$

where $E[x_i(t)]$ is the expected value of $x_i(t)$ and g_i is the link function [10].

The proposed framework was compared to the LGC for the mixture model of continuous- and discrete-valued time series and for the binary model such as spike train data.

Prior to performing simulations, a model for each time series was selected. We fitted several models with different model order to each time series and then identified the best approximating model from among a set of candidates using Akaike's information criterion (AIC) [12].

5.1. Gaussian case

As mentioned previously, when the distribution of time series is assumed to be Gaussian, the Granger causality measure based on the likelihood ratio statistic is equivalent to the LGC [6]; The log-likelihood ratio of (3) is reduced to the difference of the variances of prediction errors.

5.2. Mixture model

We analyzed the Granger causality for the mixture model of both continuous- and discrete-valued time series. For analysis, based on the causality network of Fig. 1 (a) generated

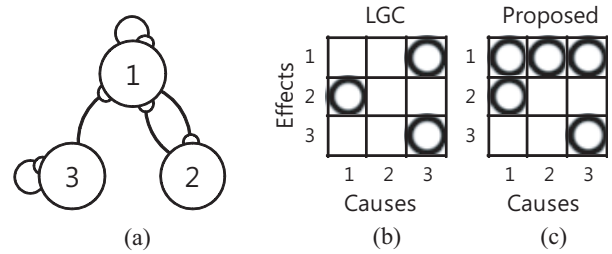


Fig. 1. (a) Three time series interact one another based on the shown network. (b) The LGC failed to estimate the original network. (c) The connectivity pattern estimated by the proposed framework exactly matched the original one of (a).

were 15,000 samples of x_1 and x_3 of Gamma distribution and x_2 of Poisson distribution, whose means depended on the past values through the link function of each distribution as (8), which are given by

$$\begin{aligned} 1/\mu_1(t|\mathbf{x}(t)) &= 1 + 0.5x_1(t-1) + x_2(t-1) + 0.5x_3(t-2), \\ \log \mu_2(t|\mathbf{x}(t)) &= 1 + x_1(t-2), \\ 1/\mu_3(t|\mathbf{x}(t)) &= 2 + 0.5x_3(t-1). \end{aligned} \quad (9)$$

We partitioned the generated data into three complementary subsets and used the first 5,000 samples for the training and the remaining 10,000 samples for the testing. The estimated causality networks using both algorithms are shown in Fig. 1 (b) and (c); White circle represents causal relationship from 'Cause' to 'Effect'. As shown in Fig. 1 (b) the LGC failed to estimate the original causal interactions; however, the proposed predicted nature of interactions between data with a high accuracy as shown in Fig. 1 (c).

5.3. Neural spike train data

In this paper, we assessed the Granger causality between synthetically generated spike train data using the proposed framework, combined with a point process statistics. The discrete, all-or-nothing nature of a sequence of spike train together with their statistical structures suggests that a neural spike train data can be regarded as a point process [13]. To model the effect of its own and other time series past values, the logarithm of the conditional intensity function, $\lambda(t|\mathbf{x}(t))$, which completely characterizes a point process, is expressed as a linear combination of the past values. This is similar to a Poisson case of the GLM. Three spike train data were generated based on the three-neuron network of Fig. 2 (a). The firing probability of each neuron is modulated by a self-inhibitory (black) interactions in addition to the inhibitory (black) and excitatory (white) interactions. The parameter vectors for the self-inhibitory, inhibitory, and excitatory interactions are given by [-1 -0.5], [-0.5 -1], and [0.5 1], respectively. Each neuron had a spontaneous firing rate of 15 Hz, and the time resolution was 1 ms. An absolute refractory period of 1 ms was

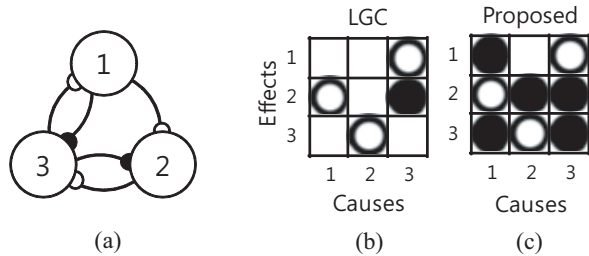


Fig. 2. (a) Three-neuron network. Each neuron was self-inhibitory. Neurons interact through inhibitory (black) and excitatory (white) connections. (b) The LGC failed to identify the original causal network. (c) The proposed framework predicted the original network with high accuracy.

enforced to prevent neurons from firing a spike in adjacent time steps. According to the above settings, 300,000 samples were generated, and the first 100,000 samples were used for training and the remaining samples used for testing.

Using the generated spike train data, we inferred the underlying causal interactions. In case of the LGC, lowpass filtering as a preprocessing step was performed to transform spike train data into continuous-valued data [2]. The inhibitory and the excitatory interactions were distinguished by the sign of the estimated parameters. The obtained results are shown in Fig. 2 (b) and (c), respectively. As shown, the causality pattern estimated by only the proposed framework exactly matched the original network of Fig. 2 (a).

6. DISCUSSION

Here we propose a probabilistic framework for assessing Granger causality between time series based on the log-likelihood ratio statistic. A Granger causality measure is proposed based on the log-likelihood ratio and a significance test to identify the original connectivity. The FDR is used to correct for multiple hypothesis testings. We show that using the Granger causality measure based on the likelihood ratio statistic the dependence measure can be decomposed into three measures. Geweke also showed the decomposition of the dependence measure in Gaussian case [6], but our study outlines a more general decomposition of the dependence measure for any distributions.

In this paper, we use the GLM framework to model time series, which can model a variety of distributions, and we show that the the proposed measure is reduced to the LGC in the Gaussian case. Besides the GLM, we can also use other distributions to assess the Granger causality between time series, and the key point is how to model the dependency of $x_i(t)$ on $\underline{x}(t)$ to fit time series well.

Simulation results showed that the proposed accurately predicted nature of interactions between discrete-valued time series as well as between continuous-valued time series. For

real data we usually don't know the ground truth, thus we used simulated data to evaluate the performance of the proposed framework in this work. In future investigations, the proposed framework will be applied to assess the causality between real data such as recorded neural spike train data.

7. REFERENCES

- [1] Brovelli et al., "Beta oscillations in a large-scale sensorimotor cortical network: directional influences revealed by Granger causality," *Proc. Natl. Acad. Sci. USA*, vol. 101, pp. 9849–9854, 2004.
- [2] Kaminski et al., "Evaluating causal relations in neural systems: Granger causality, directed transfer function and statistical assessment of significance," *Biol. Cybern.*, vol. 85, pp. 145–157, 2001.
- [3] N. Wiener, *The theory of prediction*. In: *E. F. Beckenbach (Ed) Modern Mathematics for Engineers*, McGraw-Hill, New York, 1956.
- [4] C. Granger, "Investigating causal relations by econometric models and cross-spectral methods," *Econometrica*, vol. 37, pp. 424–438, 1969.
- [5] J. Geweke, "Measures of conditional linear dependence and feedback between time series," *J. Am. Stat. Assoc.*, vol. 79, pp. 907–915, 1984.
- [6] J. Geweke, "Measurement of linear dependence and feedback between multiple time series," *J. Am. Stat. Assoc.*, vol. 77, pp. 304–313, 1982.
- [7] C. Granger, "Some recent developments in a concept of causality," *J. Econometrics*, vol. 39, pp. 199–211, 1988.
- [8] Y. Benjamini and Y. Hochberg, "Controlling the false discovery rate: a practical and powerful approach to multiple testing," *J. Roy. Stat. Soc.*, vol. 57, pp. 289–300, 1995.
- [9] Y. Pawitan, *In all likelihood: Statistical modeling and inference using likelihood*, New York: Oxford University Press, 2001.
- [10] P. McCullagh and J. Nelder, *Generalized linear models*, Chapman and Hall/CRC, 1989.
- [11] R. G. Miller, *Simultaneous statistical inference*, Springer Verlag New York, 1981.
- [12] H. Akaike, "A new look at the statistical model identification," *IEEE Transactions on Automatic Control*, vol. 19, pp. 716–723, 1974.
- [13] E. N. Brown, *Theory of point processes for neural systems*, In: *Methods and Models in Neurophysics*: Elsevier, 2005.