Quantitative Bio-Science 34(2), 63~65(2015)

Complex System and U Statistic

Jincheol Park*, Cheolyong Park, Jeongcheol Ha, Dong Hyun Yoo, Tae Yoon Kim

Department of Statistics, Keimyung University, Daegu 42601, Korea

(Received October 28, 2015; Revised November 15, 2015; Accepted November 17, 2015)

ABSTRACT

Applications of the complex system technique to various social problems have been very limited mainly due to the lack of quantitative definitions. Emergence (or scale free) is one of the most crucial concept of complex system theory and there have been many scientific and philosophical discussions without significant effort to quantify it in a practical way. In this article, quantifying the emergence in a statistical way, we investigate an interdependence amongst observations as a main player behind the emergence.

Key words: Emergence, Interdependence, U statistic

1. Introduction

Endeavoring to understand nature, physicist, chemist, and biologist have developed technologies of complex system modeling and analysis, which are highly advanced and sophisticated sufficient to explain complicated natural phenomenon. However, applications of the complex system technique developed in the natural science to various social problems arising in a daily life have been very limited. One important reason for the limitation may be a mutual lack of understanding between natural and social science: natural science is lack of understanding of social reality and social science is lack of understanding of scientific methodology. In this article, we propose a statistical framework for complex system, aiming to facilitate active research on various complex phenomena. Statistical quantification of the concepts used in the complex system theory is expected to invite active researches of various fields to investigate complex nature of problems.

Emergence (or scale free) is one of the most crucial concept of complex system theory, which is defined to be a process whereby larger entities, patterns, and regularities arise through interactions among smaller or simpler entities that themselves do not exhibit such properties [1]. Even though there have been many scientific and philosophical discussions about the concept of the emergence, it has been difficult to define it in a quantitative way. To make practical the concept of emergence, we will define it in the following quantitative way: Emergence is a phenomenon in which outliers tend to occur frequently due to a transformed distribution of observations from a normal to new χ^2. This emergence phenomenon is also called scale free in the sense that range of observation values are not limited. The scale free is oft observable for the data residing on network [2].

A possible main player behind the emergence is either interdependence amongst observations or noticeable heterogeneity of a specific group of observation. The former is called Joseph effect and the latter is called Noah effect. Considering a noticeable interdependence as a main factor for the emergence phenomenon, we will develop statistical models. In addition, it is noteworthy to investigate Noah effect from an interdependence perspective, which is reasonable because jumps in dynamic process can be explained by a stochastic process of positive association.

2. Dynamic Network and U Statistic

We propose to use the following U-statistic as a representa-

* Correspondence should be addressed to Dr. Jincheol Park, Department of Statistics, Keimyung University, Daegu 42601, Korea. Tel: +82-53-580-5189, Fax: +82-53-580-5164, E-mail: park.jincheol@gw.kmu.ac.kr

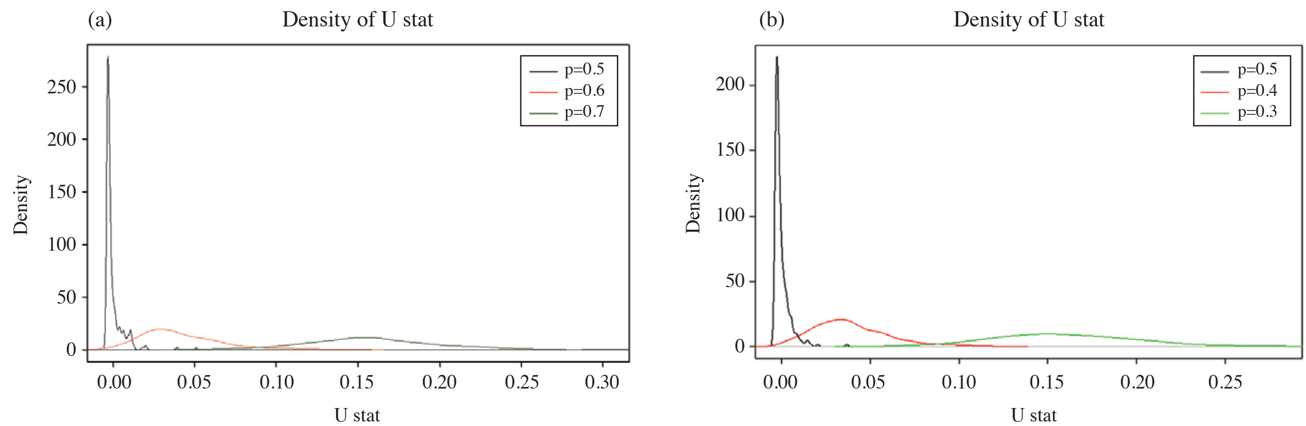

Fig. 1. Distribution of U statistic for various $p \equiv P(\varepsilon_i = 1)$. When $p = 1/2$, the distribution of U takes that of $\chi^2(1)$. However, as p gets further from 1/2, the shape of distribution gets further from that of $\chi^2(1)$ with location shift taking place. (a) Distribution of U statistic for $p = 0.5, 0.6, 0.7$; (b) Distribution of U statistic for $p = 0.5, 0.4, 0.3$.

tive quantity of a network

$$U = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1, j \neq i}^{n} \varepsilon_i \varepsilon_j,$$

where ε_i is a random variable of network node i. From the well-known facts in time series, we can see that the U statistic plays an important role in understanding dynamics of time series. For example, consider a time series $\{X_t\}$ from a network perspective between the past and the current where the time series is formulated by $AR(1)$ as $X_t = \rho X_{t-1} + e_t$. A least squares estimator of ρ for testing $\rho = 1$ is given by

$$\hat{\rho} - 1 = \sum_{i=2}^{n+1} X_{i-1} \varepsilon_i / \sum_{i=2}^{n+1} X_{i-1}^2. \tag{1}$$

Then it is easily shown that $\hat{\rho} - 1$ is U statistic because $\sum_{i=2}^{n+1} X_{i-1} \varepsilon_i$ can be rewritten as

$$\sum_{i=2}^{n+1} (\varepsilon_1 + \ldots \varepsilon_{i-1}) \varepsilon_i = \sum \sum_{i<j}^{n+1} \varepsilon_i \varepsilon_j. \tag{2}$$

As is well known, $\hat{\rho} - 1$ is called a Dickey-Fuller statistic [3], a measure for deciding whether a time series is a random walk or not so that the U statistic provides an important information on the dynamics of time series.

From the theory of U-statistic, we can obtain the following two very useful properties in analyzing network:

1. Non-degenerate U statistic ($E\{\varepsilon_i\} \neq 0$) is distributed as a normal (random network). In contrast, a degenerate U statistic ($E\{\varepsilon_i\} = 0$) is distributed as $\chi^2(1)$, which is an emergent network distribution.
2. When the interdependences amongst ε_i s get stronger, the larger location shift occurs of the distribution for U.

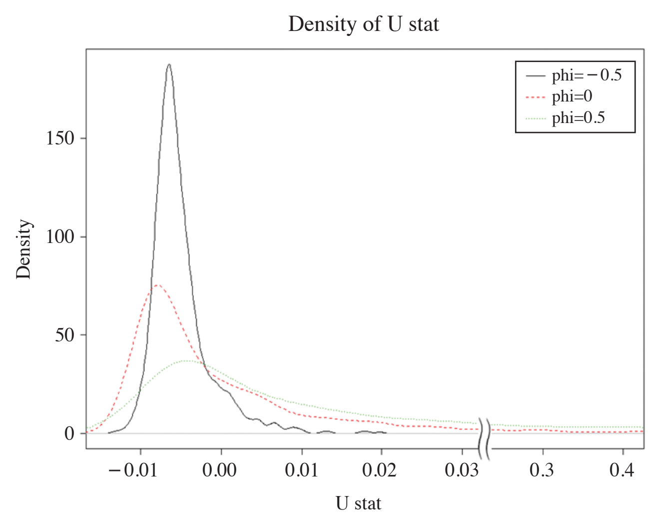

Fig. 2. The distribution of U statistic under AR(1) where $\varepsilon_t = \phi \varepsilon_{t-1} + u_t$ and u_t:$iidN\{0, (1-\phi^2)\}$. It is remarkable that as the correlation ϕ between ε_t and ε_{t-1} gets larger, location of the distribution makes more shift to right with longer tail.

To demonstrate P1 via simulation, consider independent binary random variables $\{\varepsilon_i\}$ where ε_i takes 1 or -1. Now that $E(\varepsilon_i) = 2p - 1$ where $p \equiv P(\varepsilon_i = 1)$, U can either be degenerate or non-degenerate depending upon p: $p = 1/2$ produces a degenerate statistic U distributed as $\chi^2(1)$ but for $p \neq 1/2$ a non-degenerate U statistic is yielded. As p gets further from 1/2, the distribution of U follows the more of normal distribution. See Fig. 1(a) and Fig. 1(b).

The property P2 can be widely used in investigating Joseph effect (Fig. 2). To date, researches on network has been successful in showing that the distribution of an observed network follows a χ^2 distribution, but there has been a lack of tools of analysis, prediction, and control. Therefore, a practical network complexity measure such as the suggested U-statistic enables to predict and even to control network dynamics.

3. Discussion

For effective researches and wide applications of complex system, practical quantification is necessary of complex measures. We have proposed in this article a statistical quantification of emergence, a major concept in complex system analysis, using an well-known U statistic. Future work will focus on investigation of applicability of the developed U-statistic-based statistical framework of complex system in various application fields.

References

1. O'Connor T, Wong HY. Emergent properties. In Zalta EN editor. The Stanford encyclopedia of philosophy. Summer 2015 ed. URL=http://plato.stanford.edu/entries/properties-emergent/.
2. Barabasi AL, Albert R. Emergence of scaling random networks. Science 1999;286:509-512.
3. Dickey DA, Fuller WA. Distribution of the estimators for autoregressive time series with a unit root. J Am Stat Assoc 1979; 74:427-431.