Study demonstrates a new recurrence-based method that mimics Kolmogorov-Smirnov test

August 29 2018

The recurrence plot is a vital tool for analyzing nonlinear dynamic systems, especially systems involving empirically observed time series data. RPs show patterns in a phase space system and indicate where data visit the same coordinates. RPs can also mimic some types of inferential statistics and linear analyses, such as spectral analysis. A new paper in the journal *Chaos*, provides a proof of concept for using RPs to mimic the Kolmogorov-Smirnov test, which scientists use to determine if two data sets significantly differ.

The authors, however, caution that not all types of data can be used with this new method. "Continuous data at an interval or ratio-scale level would be best suited for this technique," said Giuseppe Leonardi, one of the study's authors. "However, discretely distributed data at the same level of measurement such as dice throws would also be suitable."

The researchers analyzed recurrence points in the RPs by dividing the RP into four quadrants and counting the number of recurrence points in each cell. Then, they calculated the within-sample and between-sample recurrence rates and used those values, along with expected frequencies, to determine a p-value related to the difference between the samples. This p-value indicated whether the two groups were from the same sample or from different samples.

To verify their proof of concept, the researchers conducted a series of

simulations to see how their recurrence-based test performed compared to the Kolmogorov-Smirnov test. These simulations involved two groups of normal, skewed normal, or log-normal distributions with various combinations of means and standard deviations. The researchers found that the recurrence-based method performed roughly the same as the Kolmogorov-Smirnov test with a few differences in sensitivity with different distribution types.

The recurrence-based test appeared to be more sensitive at the tails of the distribution than the Kolmogorov-Smirnov test. This could be because the test considers deviations along the whole range of values, unlike the Kolmogorov-Smirnov test which only accounts for the largest deviation between two distributions. Leonardi explained that this enhanced sensitivity would make the recurrence-based test especially useful for nonlinear data like human reaction times.

He also cautioned that their method might suggest statistically reliable differences that are too small to be meaningful. "This might be a downside of the test for practical users," Leonardi said. "However, we have not investigated such effects in depth."

This proof of concept demonstrates that the RP can be useful for statistical analysis tools. Going forward, the team plans to investigate the effects of sample size on their method. Leonardi said they would also like to further develop the test to model other types of inferential statistics including analysis of variance.

More information: Sebastian Wallot et al, Deriving inferential statistics from recurrence plots: A recurrence-based test of differences between sample distributions and its comparison to the two-sample Kolmogorov-Smirnov test, *Chaos: An Interdisciplinary Journal of Nonlinear Science* (2018). DOI: 10.1063/1.5024915