information-dynamics-toolkit

Java Information Dynamics Toolkit for studying information-theoretic measures of computation in complex systems

[   ] Search projects

Project Home   Downloads   **Wiki**   Issues   Source   Administer

New page   Search   Current pages ▲▼   for [   ]

Search   Edit   Delete

☆ **NullDistributions**
Demonstration of analytic and null distributions of surrogates under null hypothesis for Mutual Information
examples, octave, matlab   Updated Today (moments ago) by joseph.lizier

---

Demos > Null distributions

Null distributions

This demonstration set shows how to generate bootstrap distributions of MI and TE under null hypotheses of no source-destination relationship, and also investigates the correspondence between analytic and bootstrapped distributions.

As detailed in section II.E of the paper describing the toolkit (see `InfoDynamicsToolkit.pdf` in the distribution), MI, conditional MI and TE will have non-zero bias even for unrelated data when they are estimated from a finite number of samples. Knowing the distribution of this bias is crucial for interpreting whether a sample value of TE, say, indicates a significant directed relationship between the variables, or is just an artifact of random fluctuations. One can compare a sample measurement to the distribution in order to make this judgement.

There are some known analytic forms for the distribution of these biases (see refs. 1-4 in the References section below), but otherwise the distribution must be formed by bootstrapping methods, see: Example 3 of the `Simple Java Demos`, section II.E of the toolkit paper, and refs. 5-10 below. Bootstrapping means creating surrogate source time-series, and taking a population of surrogate measurements from these surrogate source time-series to the destination. This demo compares the analytic forms (for the estimators they apply to) to those generated by bootstrapping (which is more computationally intensive, but represents the ground truth).

The demo is run under Octave or Matlab. If run in Matlab, it requires the statistics toolbox for access to the `chi2cdf` function.

The demo is available in the distribution at `demos/octave/NullDistributions`.

MI - discrete variables

We compare the analytic null distribution for MI for discrete variables against that obtained by bootstrapping in `checkMiDiscreteNullDistribution.m`. Run this for example as:

```
[mis] = checkMiDiscreteNullDistribution(10000, 100, 0.5, 0.5);
```

to examine the distribution (bootstrapped 10000 times) of MI obtained from sets of 100 samples of unrelated binary variables, each with p=0.5 probability of being a 0 or 1. A sample plot obtained from these parameters is:
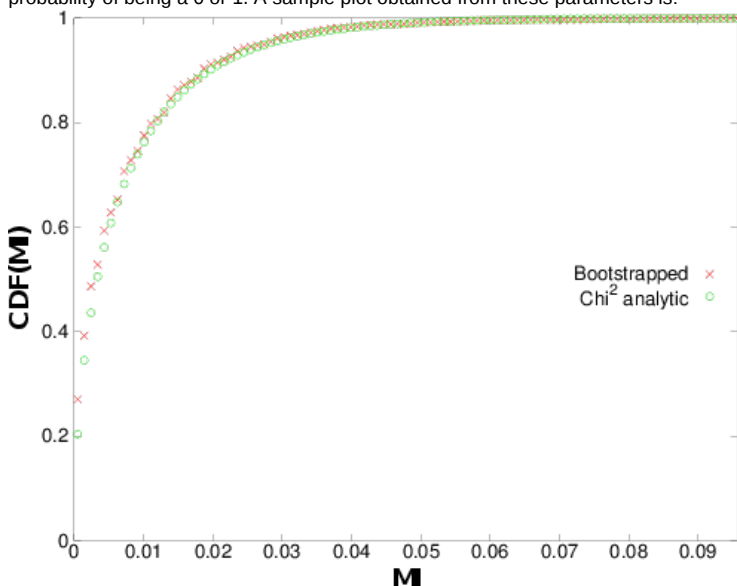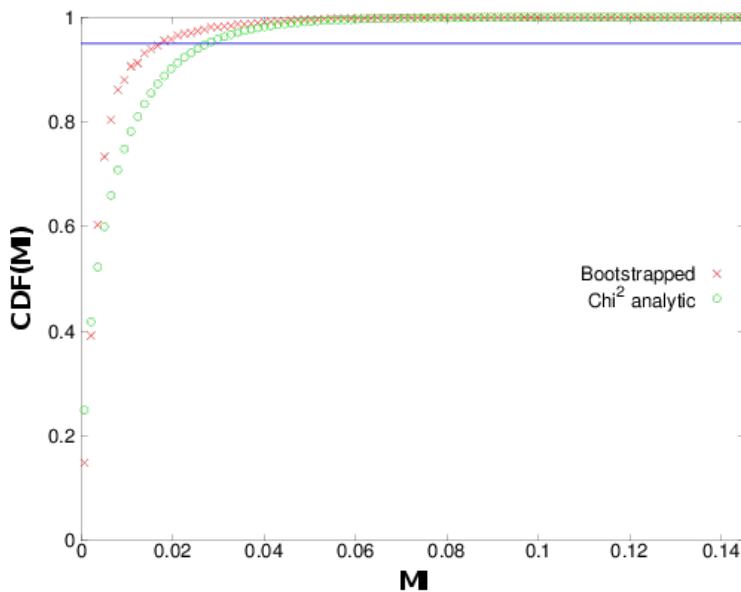

We can see the agreement between the analytic and bootstrapped versions here is very good.

The agreement is not so good for *skewed* variables however. If we run:

```
[mis] = checkMiDiscreteNullDistribution(100000, 100, 0.05, 0.05);
```

i.e. with binary variables each with only p=0.05 of being a 1, then we get the following sample plot:

In fact, the analytic distribution does not take the skewing of the variables into account (see the code), only number of samples. We know however that for skewed variables in general the MI will be lower. This is borne out in the plot above where the analytic and bootstrapped distributions do not match as well.

This can be a problem where for example we wish to use the distribution to confirm whether the sample measurement indicates a significant relationship or not. If we use an α=0.05 cutoff (i.e. CDF=0.95), represented by the blue line on the plot, we see that the analytic distribution would result in an MI threshold being around double that determined by the bootstrapping.

>> This would make a big difference to our conclusions, so for a small number of samples (which is usually what we are dealing with) **it is more accurate to stick with bootstrapping**. <<

The analytic distributions are asymptotically correct though as our number of samples gets larger. We can see this if we run our function again for the same skewed variables but for 100x more samples:

```
[mis] = checkMiDiscreteNullDistribution(100000, 10000, 0.05, 0.05);
```

which gives the following sample plot showing a good agreement now between the distributions:
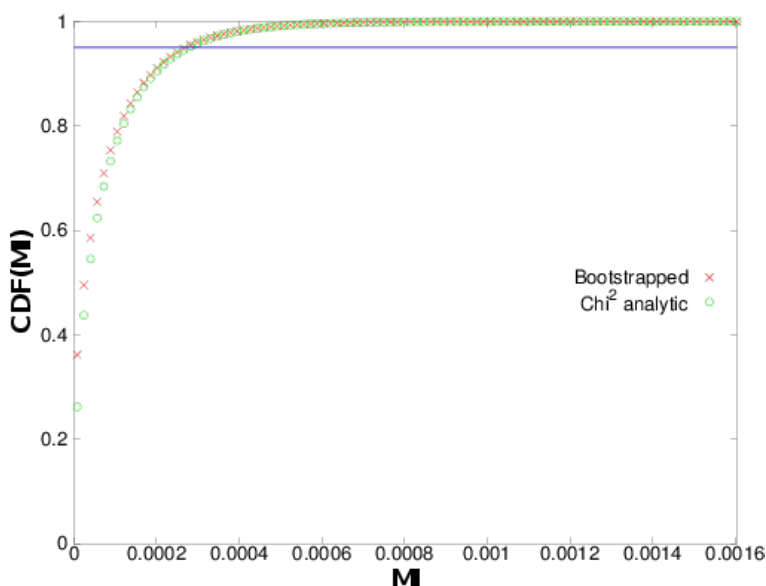
References

1. J. Geweke, Journal of the American Statistical Association **77**, 304 (1982).
2. D. R. Brillinger, Brazilian Journal of Probability and Statistics **18**, 163 (2004).
3. P. E. Cheng, J. W. Liou, M. Liou, and J. A. D. Aston, Journal of Data Science **4**, 387 (2006).
4. L. Barnett and T. Bossomaier, Physical Review Letters **109**, 138105+ (2012).
5. M. Wibral, R. Vicente, and M. Lindner, in "Directed Information Measures in Neuroscience", Understanding Complex Systems, edited by M.

   Wibral, R. Vicente, and J. T. Lizier (Springer, Berlin/Heidelberg, 2014) pp. 3-36.

6. J. T. Lizier, J. Heinzle, A. Horstmann, J.-D. Haynes, and M. Prokopenko, Journal of Computational Neuroscience **30**, 85 (2011).

7. R. Vicente, M. Wibral, M. Lindner, and G. Pipa, Journal of Computational Neuroscience **30**, 45 (2011).

8. M. Lindner, R. Vicente, V. Priesemann, and M. Wibral, BMC Neuroscience **12**, 119+ (2011).

9. M. Chávez, J. Martinerie, and M. Le Van Quyen, Journal of Neuroscience Methods **124**, 113 (2003).

10. P. F. Verdes, Physical Review E **72**, 026222+ (2005).

Enter a comment: Hint: You can use Wiki Syntax.

Submit