Fisher Information (FI)
Python Code and Tutorial

This document provides a tutorial to use the code provided to calculate the Fisher Information of any time-series dataset.

Citation: N. Ahmad, S. Derrible, T. Eason, and H. Cabezas, 2016, "Using Fisher information to track stability in multivariate systems", Royal Society Open Science, 3:160582, DOI: 10.1098/rsos.16058

Available at: http://rsos.royalsocietypublishing.org/content/3/11/160582

Table of Content **Page**

1. Pre-Requirements

1. The code has been written in Python. Thus, you will first need to install Python 2 (for v1.00) or Python 3 (for v2.00 and later), which can be downloaded freely at https://www.python.org/. Several different software packages exist that install Python along with most necessary libraries needed to run the tool, such as Anaconda: https://www.anaconda.com/download/.

2. Four codes part of this package are needed to run the tool. Make sure they are all in the same directory:
    a. fisher_main.py (main file)
    b. fisher.py
    c. smooth.py
    d. sost.py

3. Three python packages have been used to facilitate the processing and need to be installed. They are:
    a. Numpy (https://pypi.python.org/pypi/numpy)
    b. Pandas (https://pypi.python.org/pypi/pandas)
    c. Matplotlib (https://pypi.python.org/pypi/matplotlib)

4. Dataset: there should not be any missing data in the data set. If there is any missing data, zero (0) will be assigned in that missing spot. To avoid encountering problems, if a

system exists but data is missing, the choice of a value is left to the discretion of the user (e.g., using linear interpolation).

2. Required Files

Up to two different data files in .csv (comma separated value) format are required to compute Fisher Information. They are: (1) Original Data and (2) Size of State (optional). All the files must be free of any header. More about these files are given below.

Original Data

The original data, whose fisher information will be calculated, needs to be provided in .csv format. Note that all cells in the csv file must have a value, if a value is missing, zero (0) will be assigned (see pre-requirement 3). The time steps must be provided in the first column. Hence, the number of columns in the csv file must be equal to V+1, where V is the number of variables and the number of rows must be equal to the number of time steps. For example, if there are 2 (two) variables and 54 (fifty four) time steps (e.g., from 1960-2013), then a csv file with 3 columns and 54 rows is required. A sample data file named sample_data.csv is provided with this package, for which Fisher Information can be computed. As a reminder, there must be no header for any of the .csv files.

Size of State (optional)

The code provides a way to estimate the size of state automatically by taking the smallest standard deviation for each variable for a time window. However, the size of state can be an external input if preferred. Like the original data file, size of state information needs to be provided in a csv file and named as sost.csv. The number of columns for the csv file must be equal to the number of variables or number of columns of the original data file and number of rows must be one.

3. Procedure and Numerical Example

Any Python compiler can be used to run the code. In the absence of a compiler, the code can be run from the command prompt; the procedure is given below for Microsoft Windows and Macintosh computers.

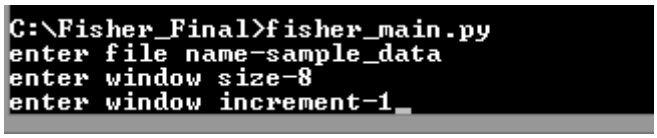
A numerical example is provided with this tutorial. Place the file “sample_data.csv” in the same folder as the code.

Step 1- Four python scripts (sost.py, fisher.py, smooth.py, fisher_main.py) are provided with this package. All the python scripts and two different csv files mentioned in the pre-requirements section need to be placed in a folder.

Step 2- Run the fisher_main.py file. The code can be run from any Python compiler or run from a command prompt (see screenshot below).

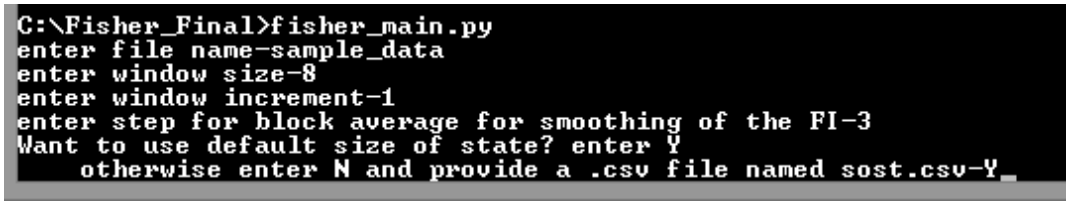
Windows: “fisher_main.py”
Mac: “python fisher_main.py”

Step 3- When asked for file name, provide the name of the original data file (e.g. sample_data) without the extension .csv. Then information about window size and window increment will be asked. It is recommended to use a window size of at least 8 and a window increment of less than the window size. For example, a window size of 8 and window increment of 1 is used here.



Step 4 – This step is to smoothen the fisher information by using a *n*-point block average, where the user will be asked to enter the value of *n* or number of points for smoothing the calculated Fisher Information. Fisher Information can be relatively volatile, capturing weekly or seasonal effects. It is therefore often desirable to smooth the values over several time steps (e.g., over 7 for 7 days in the week). For example, here three (3) is used, which means three (3) consecutive points are averaged. Note that both raw and averaged Fisher Information values are given in the output csv and plot. For example, a *n* of 3 is chosen here. The step for ticks in the X axis can also be provided by the user to avoid the text jumbling in the X axis for long data or it can left blank by simply pressing ‘Enter’, when default step of 3 will be used for xticks.

Step 5 – At this step the user will be asked whether the user wants to use a default size of sate or wants to provide a new one. If want to proceed with default size of state, needs to enter ‘Y’ and do not need to provide a size of state file. Otherwise, the user required to provide a size of state and must be named as ‘file name’_sost.csv (e.g. for here sample_data_sost.csv). For example, the default size of state is used here and ‘Y’ is entered.



Step 6- The final result will be stored in the same folder in a .csv file and named as filename_FI.csv (e.g. sample_data_FI.csv) and also the graphical representation will be stored in a pdf file as filename_FI.pdf (e.g. sample_dataFI_.pdf).

Numerical Example

Fisher Information for a sample dataset is computed here. The sample dataset has information on GDP per capita (current US$) and total population of the USA from 1960 to 2013. It therefore captures the amount of information of an unknown parameter that is present in observable dataset (i.e., GDP per capita (current US$) and total population of the USA from 1960 to 2013). Data was collected from the data bank of World Bank [1]. These two variables for each time step are used for this example. As per the requirement a csv file with three columns and 54 rows are required, where the first column must consist of time steps. The sample_data.csv is the original data file for this example. To compute the fisher information a window size of 8 and window increment of 1 is used here. Then, for number of point for block average of 3 is used here for this example. The default size of state is used for this computation. The resulted graphical representation for the sample data and time steps will be like Figure 1 and complete result is stored in the sample_data_FI.csv file. Choice of different size of state will yield different result.

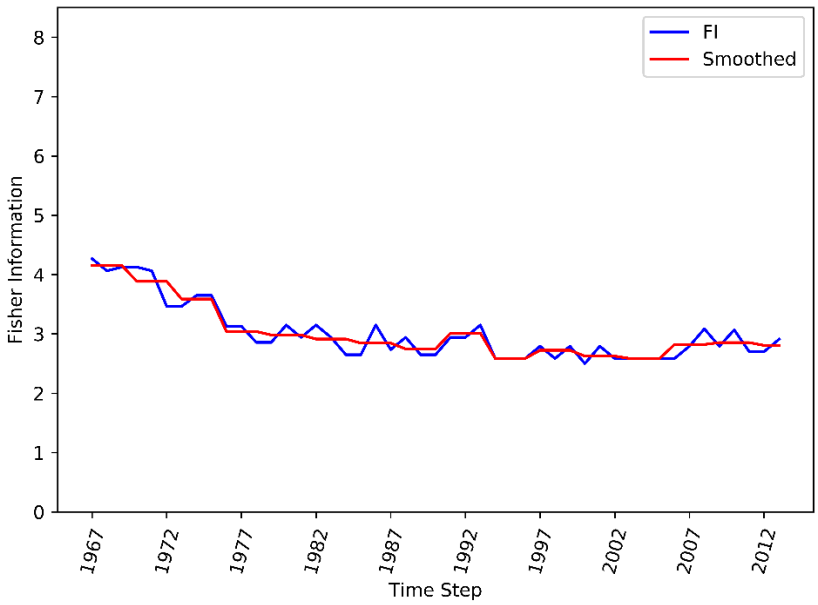


Figure 1 Fisher Information for the Sample Data

Reference

[1] "Data | The World Bank." [Online]. Available: http://data.worldbank.org/. [Accessed: 03-Apr-2015].