
correlationMatrix Documentation

Release 0.2.0

Open Risk

Feb 21, 2022

CONTENTS:

1	The correlationMatrix Library	3
1.1	Functionality	3
1.2	Architecture	3
1.3	Links to other open source software	4
2	Installation	5
2.1	Dependencies	5
2.2	From PyPi	5
2.3	From sources	5
2.4	Using virtualenv	5
2.5	File structure	6
2.6	Testing	6
3	Getting Started	7
4	Usage	9
4.1	Examples	9
4.2	Data Formats	10
5	API	11
5.1	correlationMatrix Package	11
5.2	correlationMatrix Subpackages	11
6	Roadmap	13
6.1	0.3.X	13
7	ToDo List	15
7.1	Preprocessing	15
7.2	Statistical	15
7.3	Implementation	15
8	ChangeLog	17
8.1	v0.2.0 (21-02-2022)	17
8.2	v0.1.2 (26-03-2019)	17
8.3	v0.1.0 (5-03-2019)	17
9	Indices and tables	19

NB: correlationMatrix is still in alpha release / active development. If you encounter issues please raise them in our github repository

THE CORRELATIONMATRIX LIBRARY

correlationMatrix is a Python powered library for the statistical analysis and visualization of state correlation phenomena. It can be used to analyze any dataset that captures timestamped correlations in a discrete state space. Use cases include credit rating correlations, system state event logs and more.

- Author: [Open Risk](#)
- License: Apache 2.0
- Code Documentation: [Read The Docs](#)
- Mathematical Documentation: [Open Risk Manual](#)
- Training: [Open Risk Academy](#)
- Development Website: [Github](#)
- Production Instance: [OpenCPM](#)

1.1 Functionality

You can use correlationMatrix to

- Estimate correlation matrices from historical event data using a variety of estimators
- Visualize event data and correlation matrices
- Characterise correlation matrices
- Manipulate correlation matrices (derive generators, perform comparisons, stress correlation rates etc.)
- Access standardized datasets for testing

NB: correlationMatrix is still in active development. If you encounter issues please raise them in our github repository

1.2 Architecture

- correlationMatrix supports file input/output in json and csv formats
- it has a powerful API for handling event data (based on pandas)
- provides intuitive objects for handling correlation matrices individually and as sets (based on numpy)
- supports visualization using matplotlib

1.3 Links to other open source software

- Duration based estimators are similar to etm, an R package for estimating empirical correlation matrices
- There is some overlap with lower dimensionality (survival) models like lifelines

INSTALLATION

You can install and use the correlationMatrix package in any system that supports the [Scipy ecosystem of tools](#)

2.1 Dependencies

- correlationMatrix requires Python 3
- It depends on numerical and data processing Python libraries (Numpy, Scipy, Pandas)
- The Visualization API depends on Matplotlib
- The precise dependencies are listed in the requirements.txt file.
- correlationMatrix may work with earlier versions of these packages but this has not been tested.

2.2 From PyPi

```
pip3 install pandas
pip3 install matplotlib
pip3 install correlationMatrix
```

2.3 From sources

Download the sources to your preferred directory:

```
git clone https://github.com/open-risk/correlationMatrix
```

2.4 Using virtualenv

It is advisable to install the package in a virtualenv so as not to interfere with your system's python distribution

```
virtualenv -p python3 tm_test
source tm_test/bin/activate
```

If you do not have pandas already installed make sure you install it first (will also install numpy)

```
pip3 install pandas
pip3 install matplotlib
pip3 install -r requirements.txt
```

Finally issue the install command and you are ready to go!

```
python3 setup.py install
```

2.5 File structure

The distribution has the following structure:

correlationMatrix The library source code

 model.py Main data structures

 estimators Estimator methods

 utils Helper classes and methods

examples Usage examples

datasets Contains a variety of datasets useful for getting started with correlationMatrix

tests Testing suite

2.6 Testing

It is a good idea to run the test-suite. Before you get started:

- Adjust the source directory path in correlationMatrix/__init__ and then issue the following in at the root of the distribution
- Unzip the data files in the datasets directory

```
python3 test.py
```

GETTING STARTED

Check the Usage pages in this documentation

Look at the examples directory for a variety of typical workflows.

For more in depth study, the Open Risk Academy has courses elaborating on the use of the library

- Analysis of Credit Migration using Python correlationMatrix: <https://www.openriskacademy.com/course/view.php?id=38>

The correlationMatrix packages offers a lot of functionality. Here we break down some of the main workflows for those getting started

4.1 Examples

The examples directory includes python scripts and jupyter notebooks to help you get started

- Generating correlation matrices from data
- Manipulating correlation matrices
- Visualizing correlation matrices

4.1.1 Python Scripts

Located in examples/python (For testing purposes all examples can be run using the run_examples.py script located in the root directory)

Empirical correlation Matrix

- empirical_correlation_matrix.py

Example workflows using correlationMatrix to estimate an empirical correlation matrix

Matrix Operations

- matrix_operations.py

Examples using correlationMatrix to perform various correlation matrix operations

Generate Synthetic Data

- `generate_synthetic_data.py`

Example workflows using correlationMatrix to generate synthetic data. (Edit the dataset selector to switch between examples)

Generate Visuals

- `generate_visuals.py`

Example workflows using correlationMatrix to generate visualizations

4.2 Data Formats

The correlationMatrix package supports a variety of input data formats for empirical (observation) data. Two key ones are described here in more detail. More detailed documentation about data formats provided at the correlation Matrix category at the [Open Risk Manual](#)

The correlationMatrix package structure and API.

Warning: The library is still being expanded / refactored. Significant structure and API changes are likely.

5.1 correlationMatrix Package

The core module

5.2 correlationMatrix Subpackages

5.2.1 correlationMatrix.utils subpackage

correlationMatrix.utils contents

correlationMatrix.utils Submodules

correlationMatrix.utils.converters module

correlationMatrix.utils.dataset_generators module

correlationMatrix.utils.preprocessing module

ROADMAP

`correlationMatrix` aims to become the most intuitive and versatile tool to analyse discrete correlation data. This roadmap lays out upcoming steps in this journey.

6.1 0.3.X

The 0.3.X family of releases will focus on rounding out a number of functionalities already introduced

- Stressing a set of multi-correlation matrices
- Comparing matrices produced by different estimation methods
- Further documenting the existing functionality
- Further tests, of both code and algorithms

Feature requests, bug reports and any other issues are welcome to log at the [Github Repository](#)

TODO LIST

correlationMatrix is an ongoing project. 0.1 is an alpha release

Several significant extensions are already in the pipeline. You are welcome to contribute to the development of correlationMatrix by creating Issues or Pull Requests on the github repository

7.1 Preprocessing

- More sophisticated approaches to missing data imputation

7.2 Statistical

- Further validation and characterisation of correlation matrices
- Fixing common problems encountered by empirically estimated correlation matrices
- Confidence intervals
- Additional factor models
- Network models for correlated residuals

7.3 Implementation

- PyPi installation
- Expand Sphinx documentation
- Introduce visualization objects / API
- Testing

CHANGELOG

PLEASE NOTE THAT THE API IS STILL UNSTABLE AS MORE USE CASES / FEATURES ARE ADDED REGULARLY

8.1 v0.2.0 (21-02-2022)

- **Installation:**
 - PyPI release update

8.2 v0.1.2 (26-03-2019)

- Added example `matrix_from_json_data`
- Cleaned up `PairwiseCorrelation`, `matrix_print`

8.3 v0.1.0 (5-03-2019)

- First public release of the `correlationMatrix` library

INDICES AND TABLES

- `genindex`
- `modindex`
- `search`