
icikt

Release 1.2.0

Praneeth S. Bhatt, Robert M. Flight, Hunter N.B. Moseley

Jan 31, 2023

CONTENTS:

1	icikt	1
1.1	Description	1
1.2	Installation	1
1.2.1	Install on Linux, Mac OS X	1
1.2.2	Install on Windows	1
1.2.3	Upgrade on Linux, Mac OS X	1
1.2.4	Upgrade on Windows	2
1.2.5	Get the source code	2
1.2.6	Dependencies	2
1.3	Basic usage	3
1.4	License	3
2	Documentation index:	5
2.1	icikt	5
2.1.1	Description	5
2.1.2	Installation	5
2.1.3	Basic usage	7
2.1.4	License	7
2.2	The icikt API Reference	8
2.2.1	Python Information-Content-Informed Kendall Tau Correlation (ICIKT)	8
2.3	The icikt Tutorial	9
2.3.1	Importing icikt package	9
3	Indices and tables	11
	Python Module Index	13
	Index	15

1.1 Description

The `icikt` package handles missing data before calculating a correlation between datasets for variables. The missing values are treated as information from a left-centered distribution perspective and are included in the calculation of concordant and discordant pairs used in calculating the correlation value.

Full API documentation, user guide, and tutorial can be found on [readthedocs](#)

1.2 Installation

The `icikt` package runs under Python 3.4+. Use `pip` to install. Starting with Python 3.4, `pip` is included by default.

1.2.1 Install on Linux, Mac OS X

```
python3 -m pip install icikt
```

1.2.2 Install on Windows

```
py -3 -m pip install icikt
```

1.2.3 Upgrade on Linux, Mac OS X

```
python3 -m pip install icikt --upgrade
```

1.2.4 Upgrade on Windows

```
py -3 -m pip install icikt --upgrade
```

1.2.5 Get the source code

Code is available on GitHub: <https://github.com/MoseleyBioinformaticsLab/icikt>

To clone the repo, first make sure you have `git` installed:

```
git clone https://github.com/MoseleyBioinformaticsLab/icikt.git
```

1.2.6 Dependencies

The `icikt` package depends on several Python libraries:

- `docopt` for a command line interface.
- `scipy` and `numpy` for mathematical calculations.
- `Cython` for optimized performance.

NOTE- NumPy and Cython must be preinstalled in order for this package to work.

The `pip` command will install all dependencies automatically, but if you wish to install them manually, run the following commands:

- **docopt for a command line interface**

- To install the `docopt` Python library run the following:

```
python3 -m pip install docopt # On Linux, Mac OS X
py -3 -m pip install docopt   # On Windows
```

- **scipy for performing the kendall-tau calculations**

- To install the `scipy` Python library run the following:

```
python3 -m pip install scipy # On Linux, Mac OS X
py -3 -m pip install scipy   # On Windows
```

- **numpy for creating and modifying ndarrays of data**

- To install `numpy` run the following:

```
python3 -m pip install numpy # On Linux, Mac OS X
py -3 -m pip install numpy   # On Windows
```

- **Cython for the cythonized `kendall_dis` method**

- To install the `Cython` Python library run the following:

```
python3 -m pip install Cython # On Linux, Mac OS X
py -3 -m pip install Cython   # On Windows
```

WARNING- If the following `pip` error message is generated, then the `python3` `devel` package must be installed:

```
"fatal error: Python.h: No such file or directory"
```

1.3 Basic usage

To use the `icikt` package, input a 2d array with n columns each representing an array of data for a variable. The `iciktArray` will return two $n \times n$ 2d arrays for correlations and p-values. Each element will correspond to the result of a combination of two columns in the input array. `iciktArray` can also be called from the command-line interface given the file path for the data along with several optional parameters (more in docs/tutorial).

Running through command line :

```
icikt iciktArray /path/to/file.tsv --data-format=tsv --replace=None
```

Running through python script :

```
import numpy as np
import icikt

dataArray = np.genfromtxt('/path/to/file.tsv', delimiter="\t")
# or with random values
dataArray = numpy.random.randn(100, 2)

# running just 2 arrays with icikt
corr, pVal, tMax = icikt.icikt(dataArray[:,0], dataArray[:,1])

# running all combinations with iciktArray
scaled, corrRaw, pVals, tauMax = icikt.iciktArray(dataArray)
```

1.4 License

A modified Clear BSD License

Copyright (c) 2021, Praneeth S. Bhatt, Robert M. Flight, Hunter N.B. Moseley All rights reserved.

Redistribution and use in source and binary forms, with or without modification, are permitted (subject to the limitations in the disclaimer below) provided that the following conditions are met:

- Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.
- Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.
- Neither the name of the copyright holder nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.
- If the source code is used in a published work, then proper citation of the source code must be included with the published work.

NO EXPRESS OR IMPLIED LICENSES TO ANY PARTY'S PATENT RIGHTS ARE GRANTED BY THIS LICENSE. THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN

NO EVENT SHALL THE COPYRIGHT HOLDER OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

DOCUMENTATION INDEX:

2.1 icikt

2.1.1 Description

The `icikt` package handles missing data before calculating a correlation between datasets for variables. The missing values are treated as information from a left-centered distribution perspective and are included in the calculation of concordant and discordant pairs used in calculating the correlation value.

Full API documentation, user guide, and tutorial can be found on [readthedocs](#)

2.1.2 Installation

The `icikt` package runs under Python 3.4+. Use `pip` to install. Starting with Python 3.4, `pip` is included by default.

Install on Linux, Mac OS X

```
python3 -m pip install icikt
```

Install on Windows

```
py -3 -m pip install icikt
```

Upgrade on Linux, Mac OS X

```
python3 -m pip install icikt --upgrade
```

Upgrade on Windows

```
py -3 -m pip install icikt --upgrade
```

Get the source code

Code is available on GitHub: <https://github.com/MoseleyBioinformaticsLab/icikt>

To clone the repo, first make sure you have [git](#) installed:

```
git clone https://github.com/MoseleyBioinformaticsLab/icikt.git
```

Dependencies

The **icikt** package depends on several Python libraries:

- [docopt](#) for a command line interface.
- [scipy](#) and [numpy](#) for mathematical calculations.
- [Cython](#) for optimized performance.

NOTE- NumPy and Cython must be preinstalled in order for this package to work.

The **pip** command will install all dependencies automatically, but if you wish to install them manually, run the following commands:

- **docopt for a command line interface**

- To install the docopt Python library run the following:

```
python3 -m pip install docopt # On Linux, Mac OS X
py -3 -m pip install docopt   # On Windows
```

- **scipy for performing the kendall-tau calculations**

- To install the scipy Python library run the following:

```
python3 -m pip install scipy # On Linux, Mac OS X
py -3 -m pip install scipy   # On Windows
```

- **numpy for creating and modifying ndarrays of data**

- To install numpy run the following:

```
python3 -m pip install numpy # On Linux, Mac OS X
py -3 -m pip install numpy   # On Windows
```

- **Cython for the cythonized kendall_dis method**

- To install the Cython Python library run the following:

```
python3 -m pip install Cython # On Linux, Mac OS X
py -3 -m pip install Cython   # On Windows
```

WARNING- If the following pip error message is generated, then the python3 devel package must be installed:

```
"fatal error: Python.h: No such file or directory"
```

2.1.3 Basic usage

To use the `icikt` package, input a 2d array with n columns each representing an array of data for a variable. The `iciktArray` will return two $n \times n$ 2d arrays for correlations and p-values. Each element will correspond to the result of a combination of two columns in the input array. `iciktArray` can also be called from the command-line interface given the file path for the data along with several optional parameters (more in docs/tutorial).

Running through command line :

```
icikt iciktArray /path/to/file.tsv --data-format=tsv --replace=None
```

Running through python script :

```
import numpy as np
import icikt

dataArray = np.genfromtxt('/path/to/file.tsv', delimiter="\t")
# or with random values
dataArray = numpy.random.randn(100, 2)

# running just 2 arrays with icikt
corr, pVal, tMax = icikt.icikt(dataArray[:,0], dataArray[:,1])

# running all combinations with iciktArray
scaled, corrRaw, pVals, tauMax = icikt.iciktArray(dataArray)
```

2.1.4 License

A modified Clear BSD License

Copyright (c) 2021, Praneeth S. Bhatt, Robert M. Flight, Hunter N.B. Moseley All rights reserved.

Redistribution and use in source and binary forms, with or without modification, are permitted (subject to the limitations in the disclaimer below) provided that the following conditions are met:

- Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.
- Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.
- Neither the name of the copyright holder nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.
- If the source code is used in a published work, then proper citation of the source code must be included with the published work.

NO EXPRESS OR IMPLIED LICENSES TO ANY PARTY'S PATENT RIGHTS ARE GRANTED BY THIS LICENSE. THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE COPYRIGHT HOLDER OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT

LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

2.2 The icikt API Reference

2.2.1 Python Information-Content-Informed Kendall Tau Correlation (ICIKT)

The icikt package provides a Python tool to calculate an information-content-informed Kendall Tau correlation coefficient between arrays, while also handling missing values or values which need to be removed.

`icikt.methods.icikt(x: ndarray, y: ndarray, perspective: str = 'global') → tuple`

Finds missing values, and replaces them with a value slightly smaller than the minimum between both arrays.

Parameters

- **x** – First array of data
- **y** – Second array of data
- **perspective** – perspective can be 'local' or 'global'. Default is 'global'. Global includes (NA,NA) pairs in the calculation, while local does not.

Returns

tuple with correlation, pvalue, and tauMax values

`icikt.methods.iciktArray(dataArray: ndarray, globalNA: float = 0, perspective: str = 'global', scaleMax: bool = True, diagGood: bool = True, includeOnly: Optional[tuple] = None) → tuple`

Calls iciKT to calculate ICI-Kendall-Tau between every combination of columns in the input 2d array, dataArray. Also replaces any instance of the globalNA in the array with np.nan.

Parameters

- **dataArray** – 2d array with columns of data to analyze
- **globalNA** – Optional value to replace with np.nan. Default is 0.
- **perspective** – perspective can be 'local' or 'global'. Default is 'global'. Global includes (NA,NA) pairs in the calculation, while local does not.
- **scaleMax** – should everything be scaled compared to the maximum correlation?
- **diagGood** – should the diagonal entries reflect how many entries in the sample were “good”?
- **includeOnly** – only run correlations of specified columns/combinations

Returns

tuple of the output correlations, raw correlations, pvalues, and max tau 2d arrays

Future Parameters: featureNA sampleNA

`icikt.kendall_dis.kendall_dis(x, y)`

2.3 The icikt Tutorial

2.3.1 Importing icikt package

If the icikt package is installed, it can be imported:

```
import icikt
```

Using icikt in the command-line interface

The iciktArray function can be accessed from the command line interface:

```
Either the "icikt" command or "python3 -m icikt" can be used to run the command line
↪ interface.

> icikt.py -h

Usage:
  icikt.py iciktArray <dataFilePath> [--data-format=<format>] [--replace=<globalNA>] [-
↪ -mode=<perspective>] [--scale=<scaleMax>] [--diag=<diagGood>]
  icikt.py -h | --help
```

Using a csv file with no global replace values:

```
icikt.py iciktArray test.csv --data-format=csv --replace=None
```

Using a tsv file with no global replace values:

```
icikt.py iciktArray test.tsv --data-format=tsv --replace=None
```

Using a csv file with 0 as the replace value:

```
icikt.py iciktArray test.csv --data-format=csv
```

Using a csv file in local mode with 0 as the replace value:

```
icikt.py iciktArray test.csv --mode=local --data-format=csv
```

Using icikt in a Python script

Import numpy and icikt:

```
import numpy as np
import icikt
```

Generate a numpy array from your data file:

```
dataArray = np.genfromtxt('path/to/file.tsv', delimiter='\t')
```

Call iciktArray on your dataArray, saving outputs to separate variables:

```
out, corr, pVal, tMax = icikt.iciktArray(dataArray)
print(out,corr,pVal,tMax,sep='\n\n')

# saving outputs to files
np.savetxt('outArray.csv', out, delimiter=',')
np.savetxt('corrArray.csv', corr, delimiter=',')
np.savetxt('pValArray.csv', pVal, delimiter=',')
np.savetxt('tMaxArray.csv', tMax, delimiter=',')
```

INDICES AND TABLES

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

PYTHON MODULE INDEX

i

`icikt.kendall_dis`, 8

`icikt.methods`, 8

INDEX

I

`icikt()` (*in module `icikt.methods`*), 8
`icikt.kendall_dis`
 module, 8
`icikt.methods`
 module, 8
`iciktArray()` (*in module `icikt.methods`*), 8

K

`kendall_dis()` (*in module `icikt.kendall_dis`*), 8

M

module
 `icikt.kendall_dis`, 8
 `icikt.methods`, 8