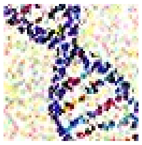
Pointillist User Manual

Bolouri Group, Institute for Systems Biology

Stephen Ramsey (sramsey at systemsbiology.org)



Pointillist Version: 1.0.4, 2006/05/08

Contents

Introduction

About Pointillist

Pointillist is a collection of programs for inferring the set elements affected by a perturbation of a biological system, based on a collection of evidences. It contains four programs: Data Manager, Data Normalizer, Significance Calculator, and Evidence-Weighted Inferer.

This document is the user manual for the Pointillist program. This manual applies to the following release version of the program:

```
release version:   1.0.4
release date:      2006/05/08
```

The README file for this version of the program can be found at the following URL:

http://magnet.systemsbiology.net/software/Pointillist/docs/ReadMe.html

The home page for this program is:

http://magnet.systemsbiology.net/software/Pointillist

If you are reading this document through a print-out, you can find the online version of this document (which may be a more recent version) at the following URL:

http://magnet.systemsbiology.net/software/Pointillist/docs/UserManual.html

A PDF version of this manual is also available on-line at:

```
http://magnet.systemsbiology.net/software/Pointillist/docs/UserManual.pdf
```

The above hyperlinks for the User Manual are for the most recent version of the Pointillist system.

External Libraries

The Pointillist system relies upon a number of external open-source libraries. These libraries are bundled with the Pointillist program and are installed within the Pointillist directory when you install Pointillist on your system.

The following table documents the external library dependencies of the Pointillist system. The libraries are provided in a compiled format called a "JAR archive". Some of the libraries have software licenses that require making the source code available, namely, the GNU Lesser General Public License (LGPL). For each of those licenses, a hyperlink is provided to a compressed archive file containing the source code for the version of the library that is included with Pointillist. These hyperlinks are shown in the "Source" column below.

Package name	JAR name	Home Page / Documentation	License	Version	Source Code
JavaHelp	`jh.jar`	`http://java.sun.com/products/javahelp`	Sun Binary Code License	1.1.3	partial
colt	`colt.jar`	`http://hoscheck.home.cern.ch/hoscheck/colt`	open source (see below)	1.0.3	full

The Colt library is provided under the following license terms:

Acknowledgements

The Pointillist software, in its current version, was implemented by Stephen Ramsey. Daehee Hwang was the architect of the statistical algorithms used within Pointillist, and provided a reference implentation of the algorithms in MATLAB. Hamid Bolouri is the Principal Investigator for this research project. Larissa Kamenkovich implemented an early prototype of a Java GUI for the Pointillist software program. William Longabaugh provided frequent advice on Java programming. The assistance, advice, and contributions of several individuals to the Pointillist project is gratefully acknowledged.

Many other individuals have contributed to the project, as well. In particular it should be noted that Pointillist makes extensive use of external libraries. The Pointillist system would not have been possible without the hard work and contributions of the authors of these libraries.

Getting Started

This section describes how to get started with using the Pointillist system.

System Requirements

The Pointillist system is implemented in the Java programming language. This means that an installation of the Java Runtime Environment (JRE) is required in order to be able to use the Pointillist system. A version of the Pointillist system installer program ("`insPoint.bin`" on Unix/Linux, or "`insPoint.exe`" on Windows) is available which has the Sun JRE version 1.4.1 pre-bundled with it. This is the recommended approach for users who are not knowledgeable in the use of the Java programming language and runtime environment.

You may also download the "thin" version of the installer that does not have the JRE pre-bundled. In order to use the "thin" installation of Pointillist, you must already have a JRE installed on your computer. The JRE must be at least version 1.4 or newer, because the software uses Java 1.4 language features and extensions. This software will not function correctly with a 1.3.X version of the JRE; if you attempt to run it under a 1.3.X version of the JRE, you will see an `UnsupportedClassVersionException`.

The specific hardware requirements for using the Pointillist system will vary depending on the complexity of the models being studied, and on the type of JRE and host operating system. A good rule of thumb is that at least 512 MB of RAM is recommended. If you are using your own JRE and it is not a Sun JRE, you will need to ensure that the appropriate command-line parameters are passed to the JRE to ensure that the built-in heap size limit is set to at least 512 MB. If you are using the Sun JRE, or the JRE that is pre-bundled with the Pointillist installer, this issue does not apply to you.

This software has been tested with the Sun Java Runtime Environment version 1.4.1 on the following platforms: Windows XP Professional on the Intel Pentium 4; Fedora Core 1 Linux on the Intel Pentium 4; Mac OSX version 10.2.6 on the PowerPC G4. It should function properly on most Windows and Linux distributions. For other operating systems, you may download the "Other Java-Enabled Platforms" version of the installer. A Mac OSX version of the installer is under development and will be released soon.

The Pointillist installer will install an executable for the Pointillist launcher program specifically designed for the operating system of the computer on which you are running the installer. This means that if you run the installer on a Windows computer, the Pointillist launcher that is installed will be a Windows executable. If there is a need to run Pointillist on multiple operating systems (e.g., in a dual-boot or heterogeneous network-file-system environment), Pointillist should be installed in a separate directory for each operating system. One exception applies: it is possible to install Pointillist on one operating system (e.g., Windows) and run it on a different operating system (e.g., Unix), if you are writing your own Java programs and just using the ISBJava API.

Launching Pointillist

Pointillist is launched by executing the one of the four executable programs (Data Manager, Data Normalizer, Significance Calculator, Evidence-Weighted Inferer) that were installed as a symbolic link by the installation program under the "Pointillist/bin" directory. The default location of the "Pointillist" directory depends on your operating system. If you are installing on a Windows computer, the directory will show up as a sub-menu in the "Start" menu. If you are installing on a Linux computer, a symbolic link is created (by default) in your home directory for each executable program in the Pointillist directory. Note that the installation program permits you to override the default location for the symbolic links to be created, so the symbolic links may not be in the default location on your computer, if you selected a different location in the installation process. By double-clicking on the one of the four symbolic links, the corresponding program should start up. Each of the four programs is described in a section below.

Data File Format

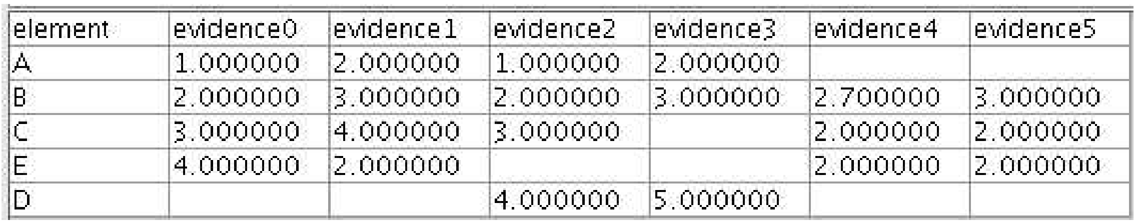
The Pointillist system is standardized on a simple text file format. A data file must be encoded in US ASCII characters, with either Unix or MS-DOS style linefeeds (either is acceptable). The file format is a matrix style, with one of three allowed delimiters: tab, space, or comma. The "space" delimiter is special in that multiple delimiter characters occurring in a row are interpreted as a single inter-cell spacing in the matrix. This allows for spaces to be used to visually align columns in the file. The fist cell of the matrix (i.e., column 1 and row 1) must contain an alphanumeric label; by convention it contains the string "element", but it can contain any simple alphanumeric string that does not contain the delimiter character for the file. The rest of the columns of the first row contain alphanumeric identifiers for the different evidence types, e.g., "evidence1", "evidence2", etc. (without quotes). Evidence names *must* be unique, i.e., you may not have two columns whose names are "evidence4". The rest of the rows of the first column contain the network element names. The Data Manager program requires the element names to be unique in some cases, but the other three programs do not require uniqueness of the network element names. However, it is a good idea to use unique network element names. If your data file has redundant network element names, perhaps you should average over the multiple observations for a given network element, to produce a file of consensus observations, before using the Pointillist program. Note that element names and evidence names must not contain the character that is used as the delimiter for the data file. If your data file is comma-delimited, your element names and evidence names must not contain the comma character.

The rest of the cells in the data file are either numeric or empty. For files using the tab and comma delimiters, empty cells may be denoted by either two delimiter characters appearing adjacent to one another (implying a "skipped" cell), or a cell containing the string "null", without the double quotes. Note that the "null" string is case sensitive, which means that you may not use the "Null" or "NULL" string to denote an empty cell. A non-empty cell must contain a floating-point number in either valid scientific notation (e.g., `1.23745E+3` or `1.23745e+3`) or in decimal notation. The number must parse successfully as a `Double` type as defined by the `java.lang.Double` class of the Java programming language.

Here is an example data file, with a comma delimiter:

```
element,evidence0,evidence1,evidence2,evidence3,evidence4,evidence5
A,1.000000,2.000000,1.000000,2.000000,,
B,2.000000,3.000000,2.000000,3.000000,2.700000,3.000000
C,3.000000,4.000000,3.000000,,2.000000,2.000000
E,4.000000,2.000000,,,2.000000,2.000000
D,,,4.000000,5.000000,,
```

which would display like this in the Data Manager:



element	evidence0	evidence1	evidence2	evidence3	evidence4	evidence5
A	1.000000	2.000000	1.000000	2.000000		
B	2.000000	3.000000	2.000000	3.000000	2.700000	3.000000
C	3.000000	4.000000	3.000000		2.000000	2.000000
E	4.000000	2.000000			2.000000	2.000000
D			4.000000	5.000000		

For files of significance values as used by the Evidence-Weighted Inferer, empty cells are denoted by two adjacent delimiters (except in the case of the "space" delimiter), or a significance value of "-1" or the "null" string.

Negative Control Data File Format

In the **Significance Calculator**, negative control data can be supplied to enable the calculation of significance values of observations from the perturbed system. The negative control data are provided in a file format that differs from the above standard Pointillist data file format. Specifically, the leftmost column of element names is *not* allowed. Although each column can have a different number of elements, *missing* elements are not allowed (i.e., each column must be a consecutive list of numbers, without skipping any rows). An example negative control data file is shown here:

```
evidence0,evidence1,evidence2,evidence3,evidence4,evidence5
1.000000,2.000000,1.000000,2.000000,3.000000,4.000000
2.000000,3.000000,2.000000,3.000000,2.700000,3.000000
3.000000,4.000000,3.000000,1.000000,2.000000,2.000000
,,4.000000,5.000000,,
```

The first row of evidence names is optional. The number of columns in the negative control data file, and the order of the columns, does not need to correspond to the data file loaded into the Significance Calculator. The Significance Calculator will prompt the user to indicate which column in the negative control data file, corresponds to the specific column selected in the loaded data matrix.

Data Manager

The **Data Manager** program enables merging of data from multiple data files that contain observations for different, partially overlapping collections of network elements. In addition, the Data Manager allows averaging over multiple measurements of the same evidence type for the same network element (whether across multiple files or within a single data file). The Data Manager displays the data loaded from one or more data files in a data table, with each row corresponding to a single element and each column corresponding to a single evidence type. The table may be sorted by evidence or element name. Either the entire table, or specific selected columns, may be saved to a data file. The data file of observations that is loaded into the Data Manager must conform to a specific format. In addition to the format requirements listed above, the Data Manager may require that element names be unique within the input data file, depending on the state of the "average duplicates" check-box (see below).

Before you load your data file, please make sure you select the proper delimiter type, using the "delimiter" drop-down list. If you attempt to load a data file with the wrong delimiter type specified, the Data Manager will likely give an error message, or (less likely) it will display incorrect data in the data table. Also, please make sure your data file conforms to the format specified above. If it does not, the Data Manager will not be able to load the data file, or it will (less likely) display incorrect data in the data table. Once you have loaded the data, it will appear in the data table. You may click on the "load data" button at any time, to load additional data files. You may change the ordering of the data files, which will rearrange the order of colums in the data table (if you have not specified a sort order for evidence names), using the "move file up" button. The "average duplicates" checkbox allows you to specify how you wish for the Data Manager to handle the case of multiple measurements of the same evidence type and network element. If you have checked this box, the Data Manager will average over the multiple measurements, and use the single resulting value as the consensus measurement. If you have not checked the box, the Data Manager will display separate rows for two elements appear with the same element name, in the same data file. In this case, if you attempt to add a second data file and the "average duplicates" box is unchecked, you will get an error message. If the first data file does not contain any duplicate elements but you attempt to add a second data file that contains an element/evidence pair that overlaps with an entry in the first data file, you will get an error message.

The following list describes the various components and controls of the Data Normalizer program window:

Load data

This button loads a file containing observations into the Input Data Table. The data file must conform to a specific format. More than one data file may be loaded into the Input Data Table, in which case the data from the multiple files will be merged. It should be noted that if two files each contain an observation for element X and evidence type Y (or if a single file contains an element X that is represented on more than one row), the "Average duplicates" check-box should be selected, or else an error message will appear when the data file(s) are loaded. This is because the Data Manager needs to average redundant data before it can merge data files that have overlapping observations.

The "Average duplicates" check-box gives the Data Manager permission to carry out this averaging.

Delimiter

This drop-down list is used to specify how your input data file (and output data file) are delimited. The three choices are "tab", "space", and "comma". Comma-delimited files are recommended for reasons of unambiguity and compatibility with spreadsheet programs. If you choose "space", be advised that there is a restriction on how you can specify an empty cell (missing data). With a "space" delimiter, an empty cell must be denoted by the string "null" (without double quotes). Note that the string is case-sensitive, so you cannot use the string "Null" or "NULL". Once you have loaded your input data file, you may change the delimiter that you use to save the output file, if you wish.

Average duplicates

This check-box is used to indicate how the Data Manager should handle overlapping ovservations in the input data. Overlapping observations are two observations for the same element name and evidence type. If this check-box is selected, the Data Manager will average over any overlapping observations to assign a "consensus" observation to the evidence type and element. If this check-box is not selected, the Data Manager will display separate rows for two elements appear with the same element name, in the same data file. In this case, if you attempt to add a second data file and the "average duplicates" box is unchecked, you will get an error message. If the first data file does not contain any duplicate elements but you attempt to add a second data file that contains an element/evidence pair that overlaps with an entry in the first data file, you will get an error message. Note that two data files may also contain *complementary* data, i.e., data for the same evidence types but different elements; this is not overlapping data, and is not impacted by the "allow duplicates" check-box. The default is for this check-box to be selected.

Input data table

This table displays the observations that have been loaded into the program, in the large table in the middle of the program window. The first row contains the evidence types, and the first column contains the network element names. All other cells are either numeric values (denoting the results of observations), or are empty. All cells may be edited, subject to the constraint that the file format specification is not violated (i.e., you cannot enter text into the numeric data cells, change an element name to become empty, etc.). A column of data may be selected by clicking the mouse on one of the columns of the table. By holding down the control key while clicking on multiple columns in succession, more than one column may be jointly selected. Column selection is used to designate those columns that will be saved to a file, if the "save selected columns" button is pressed.

Element sort status / Evidence sort status

These two drop-down lists are used to sort the rows and columns of the input data table, respectively.

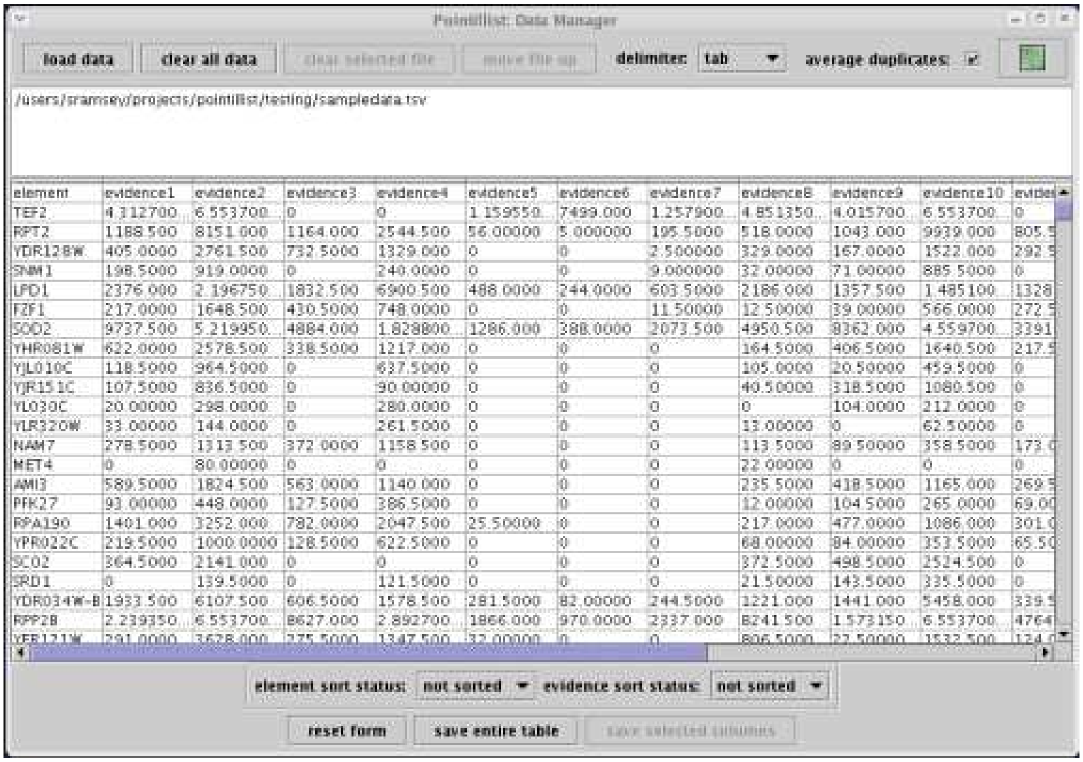
Save entire table

This button saves the entire contents of the Input Data Table to a file. The delimiter used for the file is as specified in the "Delimiter" drop-down list at the top of the program window. When the input data table is saved to a file, the ordering of the rows and columns will be as it appears in the input data table. An empty cell will be denoted by two adjacent delimiter characters, except in the case of the "space" delimiter, where the "null" string will instead be used to denote an empty cell.

Save selected columns

This button saves the selected columns of the Input Data Table to a file. The delimiter used for the file is as specified in the "Delimiter" drop-down list at the top of the program window. When the input data table is saved to a file, the ordering of the rows and columns will be as it appears in the input data table. An empty cell will be denoted by two adjacent delimiter characters, except in the case of the "space" delimiter, where the "null" string will instead be used to denote an empty cell. This button is only enabled if at least one column of the input data table is selected.

Here is a screen shot of the Data Manager program:



delimiter: tab

average duplicates: ☑

/users/sramsey/projects/pointillist/testing/sampledata.tsv

element	evidence1	evidence2	evidence3	evidence4	evidence5	evidence6	evidence7	evidence8	evidence9	evidence10	evide
TEF2	4.312700...	6.553700...	0	0	1.159550...	7499.000	1.257900...	4.851350...	4.015700...	6.553700...	0
RPT2	1188.500	8151.000	1164.000	2544.500	56.00000	5.000000	195.5000	518.0000	1043.000	9939.000	805.5
YDR128W	405.0000	2761.500	732.5000	1329.000	0	0	2.500000	329.0000	167.0000	1522.000	292.5
SNM1	198.5000	919.0000	0	240.0000	0	0	9.000000	32.00000	71.00000	885.5000	0
LPD1	2376.000	2.196750...	1832.500	6900.500	488.0000	244.0000	603.5000	2186.000	1357.500	1.485100...	1328
FZF1	217.0000	1648.500	430.5000	748.0000	0	0	11.50000	12.50000	39.00000	566.0000	272.5
SOD2	9737.500	5.219950...	4884.000	1.828800...	1286.000	388.0000	2073.500	4950.500	8362.000	4.559700...	3391
YHR081W	622.0000	2578.500	338.5000	1217.000	0	0	0	164.5000	406.5000	1640.500	217.5
YJL010C	118.5000	964.5000	0	637.5000	0	0	0	105.0000	20.50000	459.5000	0
YJR151C	107.5000	836.5000	0	90.00000	0	0	0	40.50000	318.5000	1080.500	0
YL030C	20.00000	298.0000	0	280.0000	0	0	0	0	104.0000	212.0000	0
YLR320W	33.00000	144.0000	0	261.5000	0	0	0	13.00000	0	62.50000	0
NAM7	278.5000	1313.500	372.0000	1158.500	0	0	0	113.5000	89.50000	358.5000	173.0
MET4	0	80.00000	0	0	0	0	0	22.00000	0	0	0
AMI3	589.5000	1824.500	563.0000	1140.000	0	0	0	235.5000	418.5000	1165.000	269.5
PFK27	93.00000	448.0000	127.5000	386.5000	0	0	0	12.00000	104.5000	265.0000	69.00
RPA190	1401.000	3252.000	782.0000	2047.500	25.50000	0	0	217.0000	477.0000	1086.000	301.0
YPR022C	219.5000	1000.0000	128.5000	622.5000	0	0	0	68.00000	84.00000	353.5000	65.50
SCO2	364.5000	2141.000	0	0	0	0	0	372.5000	498.5000	2524.500	0
SRD1	0	139.5000	0	121.5000	0	0	0	21.50000	143.5000	335.5000	0
YDR034W-B	1933.500	6107.500	606.5000	1578.500	281.5000	82.00000	244.5000	1221.000	1441.000	5458.000	339.5
RPP2B	2.239350...	6.553700...	8627.000	2.892700...	1866.000	970.0000	2337.000	8241.500	1.573150...	6.553700...	4764
YFR121W	291.0000	3678.000	275.5000	1347.500	32.00000	0	0	806.5000	22.50000	1532.500	124.0

element sort status: not sorted

evidence sort status: not sorted

Data Normalizer

The **Data Normalizer** is a program that can be used to perform a normalization of microarray expression data that is arranged in a matrix format in a single data file. The data file must conform to the a specific format. Currently, the only normalization method supported is Quantile Normalization. The quantile normalization algorithm implemented here is based on a prototype written by Daehee Hwang at Institute for Systems Biology, and it is similar to the quantile normalization algorithm proposed by Bolstat *et al.* in their paper

> Bolstad, B.M., Irizarry R. A., Astrand M., and Speed, T.P. (2003), "A Comparison of Normalization Methods for High Density Oligonucleotide Array Data Based on Bias and Variance." *Bioinformatics* **19**(2):185-193

Note that only the quantile normalization step of the RMA (Robust Multi-Chi Average) procedure is implemented in this class; background adjustment is not implemented here, and is assumed to have been applied to the raw observations before this program is applied to the data. Each column of the data file corresponds to a different microarray experiment, and each row of the data file corresponds to a different probe. The first row must contain experiment labels, and each column's experiment label must be unique. The first column must contain probe names, although there is no uniqueness requirement for probe names. The cell in the first row and the first column should contain an arbitrary string identifier such as "element", that does not contain the file delimiter (the actual string used is not important; the cell is just a placeholder). Cells in the data file may be empty, just as described in the Data Manager section above.

The following list describes the various components and controls of the Data Normalizer program window:

Load observations

This button is used to load a file of data into the program. The data is displayed in the input data table in the top half of the program window. Only one file may be loaded into the input data table at a time.

Delimiter

This drop-down list is used to specify how your input data file (and output data file) are delimited. The three choices are "tab", "space", and "comma". Comma-delimited files are recommended for reasons of unambiguity and compatibility with spreadsheet programs. If you choose "space", be advised that there is a restriction on how you can specify an empty cell (missing data). With a "space" delimiter, an empty cell must be denoted by the string "null" (without double quotes). Note that the string is case-sensitive, so you cannot use the string "Null" or "NULL". Once you have loaded your input data file, you may change the delimiter that you use to save the output file, if you wish.

Input data table

This table displays the observations that have been loaded into the program, in the top third of the program window. The column header names are evidence types, and the first column contains the network element names. All other cells are either numeric values (denoting the results of observations), or are empty.

Normalize using the scale

This drop-down list is used to specify how the data is to be scaled. The options are "norm_only" and "logarithm". If you specify "norm_only", only a normalization is performed, and the data is not rescaled. If you specify "logarithm", the logarithm of each observation is taken before the normalization is performed. In addition, the "logarithm" option automatically selects the "Fix negative values" check-box (see below), to ensure that there are no negative observation values when the logarithm is taken. The "logarithm" option is typically used for normalizing Affymetrix microarray data.

Specify the error tolerance

This type-in box is used to specify the error tolerance for estimating missing data values. If there are no missing data in the file of observations loaded into the program, this type-in box will be left disabled. When there are missing data values, this type-in box will be enabled and pre-populated with a small positive floating-point value. This value controls the threshold for the average fractional error in estimating the missing data values. The average fractional error is computed as the average over all missing observations, of the fractional deviation in the interpolated missing observation from one iteration to the next. Missing observations are estimated as the median of all non-missing observations in the column (evidence type). The missing observations thus estimated are used only in computing quantiles for normalizing the non-missing data; the missing observations will be removed when the normalized observations are written out by the program. If you have a really small data set, it is best to set this error tolerance to be a bit larger, perhaps 0.2. For large data sets (more than a few hundred elements), the default value should work. The error tolerance must be a positive floating-point number, and less than 1.0.

Fix negative values

This check-box is used to specify whether the data should be additively corrected to ensure that no observations are negative, before the rescaling is applied. Choosing the "logarithm" option for the normalization scale will automatically select this check-box. If you leave the box checked, the data will be additively corrected. If you de-select the check-box, the data will not be additively corrected. Note that if you have nonpositive data values and you selected "logarithm" for the scale and left this box un-checked, you would get an error message when you attempt to normalize the data. If all your data is positive, nothing will be changed by this option. The "norm_only" scale option will not give an error message if there are nonpositive observations, and the normalization will be successful in that case.

Max iterations

This type-in box is used to specify the maximum number of iterations that are used for estimating missing data values. If there is no missing data, this type-in box is left disabled. If there are empty cells in the data table, this type-in box will be enabled, and will be pre-populated with a small value. You may change it to any positive integer value. This value places a hard limit on the number of iterations of the normalization algorithm, regardless of the error threshold you selected (see "Specify the error tolerance" above). If you have an extremely large data set, you may wish to select a small number of iterations, perhaps in the range 3-5. Alternatively, you could specify a more permissive error tolerance. Note that if you specify an extremely small error tolerance and an extremely large maximum number of iterations, the normalization algorithm could run for a very long time. If the algorithm fails to complete in a timely fashion, you can stop it by closing the program's window using the window manager for your computer.

Normalize raw observations

This button starts the normalization. The normalized observations are displayed in a table in the lower half of the program window.

Iteration count

This box displays the number of iterations required before the normalization process converged to the specified error tolerance. If no cells were empty (denoting missing data) in the input data table, the number of iterations will always be 1. If there are empty cells in the input data table, the number of iterations may be greater than 1.

Fractional error

This box displays the average fractional error, which is only applicable when there are empty cells in the input data table. If there are empty cells in the input data table, the average fractional error is computed as the average of the fractional change in the estimated values of the "empty cells" from one iteration to the next. The value displayed in this box may or may not be less than the error threshold specified above, depending on the value chosen for the "max iterations" field.

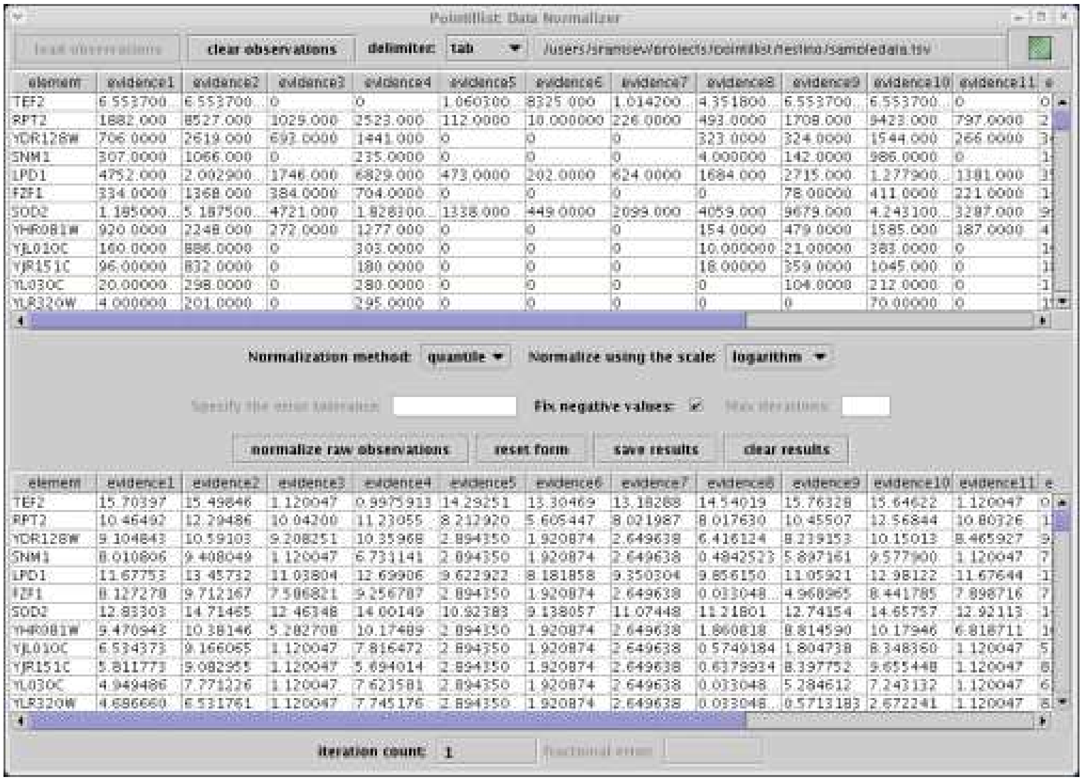
Results Table

This table displays the results of the data normalization. The columns should have approximately the same quantiles, although they will not be exactly equal if there is missing data.

Save results

This button opens a file dialog to specify the file to which the results should be saved. The delimiter specified in the "Delimiter:" drop-down list is used. Results are saved in a matrix style, in a text file, with columns representing evidence types and rows representing elements. If the delimiter type is "comma" or "tab", a missing observation is denoted by an empty cell; if the delimiter type is "space", a missing observation is denoted by "null". If you change the delimiter type between when you load the data file and when you save it, and if some element or evidence names contain the new delimiter character, the delimiter characters appearing in the evidence/element names will be converted to underscore ("_") characters.

Here is a screen shot of the Data Normalizer program:



Pointillist: Data Normalizer
load observations
clear observations
delimiter: tab
/users/sramsey/projects/pointillist/testing/sampledata.tsv

element	evidence1	evidence2	evidence3	evidence4	evidence5	evidence6	evidence7	evidence8	evidence9	evidence10	evidence11	e
TEF2	6.553700...	6.553700...	0	0	1.060300...	8325.000	1.014200...	4.351800...	6.553700...	6.553700...	0	0
RPT2	1882.000	8527.000	1029.000	2523.000	112.0000	10.000000	226.0000	493.0000	1708.000	9423.000	797.0000	2
YDR128W	706.0000	2619.000	693.0000	1441.000	0	0	0	323.0000	324.0000	1544.000	266.0000	3
SNM1	307.0000	1066.000	0	235.0000	0	0	0	4.000000	142.0000	986.0000	0	1
LPD1	4752.000	2.002900...	1746.000	6829.000	473.0000	202.0000	624.0000	1684.000	2715.000	1.277900...	1381.000	3
FZF1	334.0000	1368.000	384.0000	704.0000	0	0	0	0	78.00000	411.0000	221.0000	1
SOD2	1.185000...	5.187500...	4721.000	1.828100...	1338.000	449.0000	2099.000	4059.000	9679.000	4.243100...	3287.000	9
YHR081W	920.0000	2248.000	272.0000	1277.000	0	0	0	154.0000	479.0000	1585.000	187.0000	4
YJL010C	160.0000	886.0000	0	303.0000	0	0	0	10.000000	21.00000	383.0000	0	1
YJR151C	96.00000	832.0000	0	180.0000	0	0	0	18.00000	359.0000	1045.000	0	1
YL030C	20.00000	298.0000	0	280.0000	0	0	0	0	104.0000	212.0000	0	1
YLR320W	4.000000	201.0000	0	295.0000	0	0	0	0	0	70.00000	0	1

Normalization method: quantile

Normalize using the scale: logarithm

Specify the error tolerance:

Fix negative values: ☑

Max iterations:

element	evidence1	evidence2	evidence3	evidence4	evidence5	evidence6	evidence7	evidence8	evidence9	evidence10	evidence11	e
TEF2	15.70397	15.49846	1.120047	0.9975913	14.29251	13.30469	13.18288	14.54019	15.76328	15.64622	1.120047	0
RPT2	10.46492	12.29486	10.04200	11.23055	8.212920	5.605447	8.021987	8.017630	10.45507	12.56844	10.80326	1
YDR128W	9.104843	10.59103	9.208251	10.35968	2.894350	1.920874	2.649638	6.416124	8.239153	10.15013	8.465927	9
SNM1	8.010806	9.408049	1.120047	6.731141	2.894350	1.920874	2.649638	0.4842523	5.897161	9.577900	1.120047	7
LPD1	11.67753	13.45732	11.03804	12.69906	9.622922	8.181858	9.350304	9.856150	11.05921	12.98122	11.67644	1
FZF1	8.127278	9.712167	7.586821	9.256787	2.894350	1.920874	2.649638	0.033048...	4.968965	8.441785	7.898716	7
SOD2	12.83303	14.71465	12.46348	14.00149	10.92383	9.138057	11.07448	11.21801	12.74154	14.65757	12.92113	1
YHR081W	9.470943	10.38146	5.282708	10.17489	2.894350	1.920874	2.649638	1.860818	8.814590	10.17946	6.818711	1
YJL010C	6.534373	9.166065	1.120047	7.816472	2.894350	1.920874	2.649638	0.5749184	1.804738	8.348360	1.120047	5
YJR151C	5.811773	9.082955	1.120047	5.694014	2.894350	1.920874	2.649638	0.6379934	8.397752	9.655448	1.120047	8
YL030C	4.949486	7.771226	1.120047	7.623581	2.894350	1.920874	2.649638	0.033048	5.284612	7.243132	1.120047	6
YLR320W	4.686660	6.531761	1.120047	7.745176	2.894350	1.920874	2.649638	0.033048...	0.5713183	2.672241	1.120047	8

iteration count: 1

fractional error:

Significance Calculator

The **Significance Calculator** is a program that can analyze the probability distribution of a set of observations, and compute the statistical signficance of each observation on the basis of the distribution. Alternatively, the significances can be computed based on the distribution of a separate set of "negative control" observations. The "significance" of an observation is here defined as the probability that the observation would occur by chance, given either the global or the "negative control" distribution for observations. The input to the program is a matrix of observations, which must conform to a specific format. Missing observations are allowed, and are denoted by an empty cell or the string "null" (quotes not included, case sensitive).

The following list describes the various components and controls of the Significance Calculator program window:

Load observations
: This button loads a file of observations. The observations are displayed in the Input Data Table, in the upper third of the program window.. Only one file may be loaded into the input data table at a time.

Delimiter
: This drop-down list is used to specify how your input data file (and output data file) are delimited. The three choices are "tab", "space", and "comma". Comma-delimited files are recommended for reasons of unambiguity and compatibility with spreadsheet programs. If you choose "space", be advised that there is a restriction on how you can specify an empty cell (missing data). With a "space" delimiter, an empty cell must be denoted by the string "null" (without double quotes). Note that the

string is case-sensitive, so you cannot use the string "Null" or "NULL". Once you have loaded your input data file, you may change the delimiter that you use to save the output file, if you wish.

Input data table

This table displays the observations that have been loaded into the program, in the top third of the program window. The column header names are evidence types, and the first column contains the network element names. All other cells are either numeric values (denoting the results of observations), or are empty.

Evidence choices table

This table displays the list of evidence types, and is activated when the input data file is loaded. The five rightmost columns of the table are used to specify how significances are to be computed for each type of evidence. Each column type is defined below.

Single-tailed

This checkbox is used to define whether the observations for a given evidence are to be converted to a significance using a single-tailed or two-tailed test. A single-tailed test is appropriate if the observations are never negative, or if the observations are always greater than some minimum value (e.g., 1.0). A two-tailed test is appropriate if observations can have large negative or large positive values, e.g., logarithm of a fold-change for protein or mRNA levels. The program attempts to guess whether the distribution is single-tailed or two-tailed, by looking to see if the evidence type has any negative observations. You can override the initial guess by clicking on the "single-tailed" checkbox.

Number of bins

This text field is used to specify the number of bins that are to be used to calculate the nonparametric distribution for the observations for the given evidence type. The program makes an initial guess at an appropriate number of bins, but you may change it to any integer value greater than 1.0.

Smoothing length

This text field is used to specify the smoothing length to be used; this is the standard deviation of the Gaussian used as a kernel density for smoothing the distribution of observations. The program makes an initial guess at an appropriate smoothing lenght, but you may change it to any value greater than 0.0.

Compute significances

This checkbox is used to specify whether significances should be computed for this evidence type. By default, the box is checked, meaning the program will compute significances for this evidence type and include it as a column in the reusults data table. If you uncheck this box, the program will not compute significances for this data type, and will omit the corresponding column from the results data table.

Negative control observations

This field is used to specify a set of "negative control" observations for the evidence type, as described above. It is optional, and by default, set to "<none>". By double-clicking on the field, a file browser is opened and you are prompted to select a data file. The data file should contain negative control data arranged in columns, with each column a different evidence type. The file can contain only one column, or multiple columns. If you specify a file with multiple columns, the program will prompt you to indicate which column is to be used as the negative control for the selected column in the loaded data matrix. The columns may have different numbers of observations in them, and there is no requirement of uniformity of rows. You will be prompted to select a column in a separate data table that is displayed, for the negative control file you provided. If you wish to delete your choice of negative control data for a given evidence, just select the "negative control" cell for that evidence, and press the "delete" key.

Calculate the significances using the method

This drop-down list is used to specify how the significances are to be computed. The same method is applied to all evidence types in the input data table. The four methods allowed are:

- **CDF_NONPARAMETRIC**
  Use a nonparametric distribution with a gaussian kernel density, to calculate the significance using the "cumulative density function" (CDF), which is the area under the probability density function (PDF). The distribution is smoothed using the gaussian kernel density with standard deviation equal to the specified "smoothing length". The CDF method is also known as the Fisher Test. Twice the area under the tail of the distribution closest to the observation is used to compute the significance.
- **PDF_NONPARAMETRIC**
  Use a nonparametric distribution with a gaussian kernel density, to calculate the significance using the "probability density function" (PDF). The distribution is smoothed using the gaussian kernel density with standard deviation equal to the specified "smoothing length". The PDF method is also called the Bayesian method.
- **CDF_PARAMETRIC**
  Use a parametric distribution to calculate the significance. The distribution is a best-fit to the data using one or two parameters. Various theoretical distributions are attempted, and the one that fits the data best is used. If no theoretical distribution fits the data within the "maximum chi square" specified, the significance calculation will fail with an error message. The significance is calculated using the area under the theoretical distribution (i.e., the CDF).
- **PDF_PARAMETRIC**
  Use a parametric distribution to calculate the significance. The distribution is a best-fit to the data using one or two parameters. Various theoretical distributions are attempted, and the one that fits the data best is used. If no theoretical distribution fits the data within the "maximum chi square" specified, the significance calculation will fail with an error message. The significance is calculated using the density of the theoretical distribution (i.e., the PDF).

If you are unsure of which method to use, it is recommended to use the default, "CDF_NONPARAMETRIC", because the default settings in the Evidence-Weighted Inferer are appropriate to a "CDF_NONPARAMETRIC" choice in the Significance Calculator.

Calculate significances

This button causes the program to calculate significances for all evidence types in the input data table for which the "compute significance" checkbox in the Evidence Choices Table is checked. The results are displayed in the Results Table.

Results Table

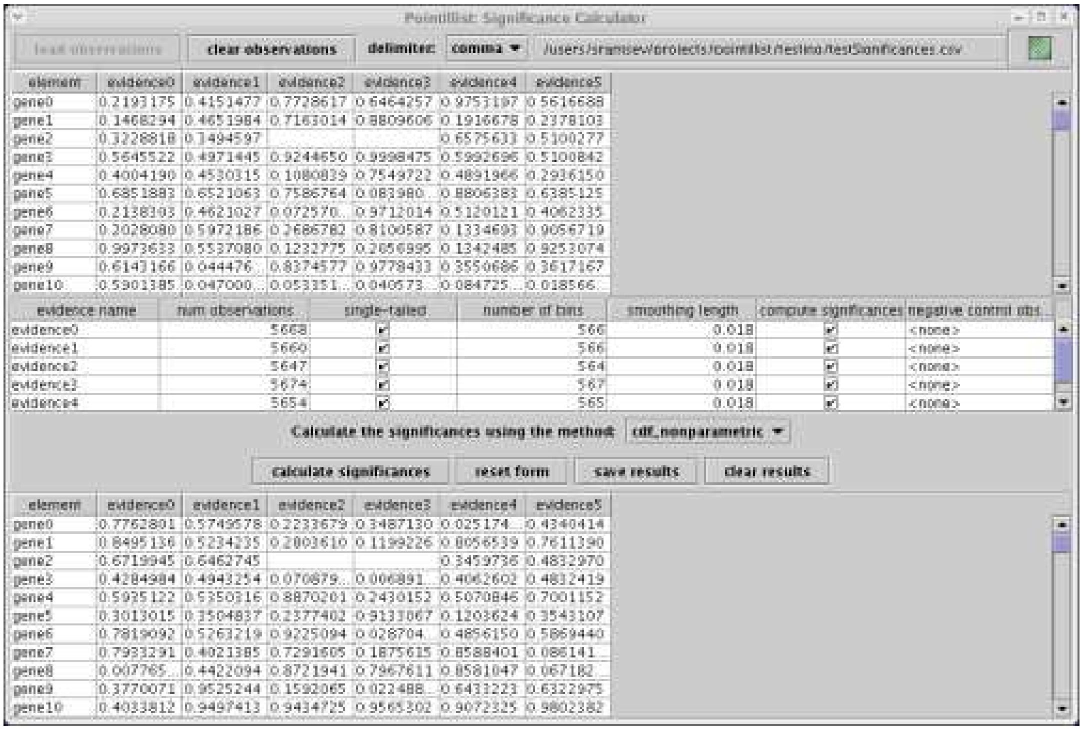
This table displays the results of the significance calculation. Smaller significance values indicate a smaller probability that a given observation could have occurred by chance, given the negative control distribution.

Save results

This button opens a file dialog to specify the file to which the results should be saved. The delimiter specified in the "Delimiter:" drop-down list is used. Results are saved in a matrix style, in a text file, with columns representing evidence types and rows representing elements. If the delimiter type is "comma" or "tab", a missing observation is denoted by an empty cell; if the delimiter type is "space", a missing observation is denoted by "-1". If you change the delimiter type between when you load the data file and when you save it, and if some element or evidence names contain the new delimiter character, the delimiter characters appearing in the evidence/element names will be converted to underscore ("_") characters.

If you have a large data set, and you are using the `CDF_NONPARAMETRIC` formula for calculating the significances, the Significance Calculator may take a long time to calculate the significances. Please be patient. On Linux, the IBM Java Virtual Machine has been found to give a substantial performance improvement, when using the `CDF_NONPARAMETRIC` formula.

Here is a screen shot of the Significance Calculator program:



delimiter: comma ▼ /users/sramsey/projects/pointillist/testing/testSignificances.csv

element	evidence0	evidence1	evidence2	evidence3	evidence4	evidence5
gene0	0.2193175	0.4151477	0.7728617	0.6464257	0.9753197	0.5616688
gene1	0.1468294	0.4651984	0.7163014	0.8809606	0.1916678	0.2378103
gene2	0.3228818	0.1494597			0.6575633	0.5100277
gene3	0.5645522	0.4971445	0.9244650	0.9998475	0.5992696	0.5100842
gene4	0.4004190	0.4530315	0.1080839	0.7549722	0.4891966	0.2936150
gene5	0.6851883	0.6521063	0.7586764	0.083980...	0.8806383	0.6385125
gene6	0.2138303	0.4621027	0.072570...	0.9712014	0.5120121	0.4062335
gene7	0.2028080	0.5972186	0.2686782	0.8100587	0.1334693	0.9056719
gene8	0.9973633	0.5537080	0.1232775	0.2656995	0.1342485	0.9253074
gene9	0.6143166	0.044476...	0.8374577	0.9778433	0.3550686	0.3617167
gene10	0.5901385	0.047000...	0.053351...	0.040573...	0.084725...	0.018566...

evidence name	num observations	single-tailed	number of bins	smoothing length	compute significances	negative control obs...
evidence0	5668	☑	566	0.018	☑	<none>
evidence1	5660	☑	566	0.018	☑	<none>
evidence2	5647	☑	564	0.018	☑	<none>
evidence3	5674	☑	567	0.018	☑	<none>
evidence4	5654	☑	565	0.018	☑	<none>

Calculate the significances using the method: cdf_nonparametric ▼

element	evidence0	evidence1	evidence2	evidence3	evidence4	evidence5
gene0	0.7762801	0.5749578	0.2233679	0.3487130	0.025174...	0.4340414
gene1	0.8495136	0.5234235	0.2803610	0.1199226	0.8056539	0.7611390
gene2	0.6719945	0.6462745			0.3459736	0.4832970
gene3	0.4284984	0.4943254	0.070879...	0.006891...	0.4062602	0.4812419
gene4	0.5925122	0.5350316	0.8870201	0.2430152	0.5070846	0.7001152
gene5	0.3013015	0.3504837	0.2377402	0.9133067	0.1203624	0.3543107
gene6	0.7819092	0.5263219	0.9225094	0.028704...	0.4856150	0.5869440
gene7	0.7933291	0.4021385	0.7291605	0.1875615	0.8588401	0.086141...
gene8	0.007765...	0.4422094	0.8721941	0.7967611	0.8581047	0.067182...
gene9	0.3770071	0.9525244	0.1592065	0.022488...	0.6433223	0.6322975
gene10	0.4033812	0.9497413	0.9434725	0.9565302	0.9072325	0.9802382

Evidence-Weighted Inferer

The **Evidence-Weighted Inferer** is a classification program that attempts to divide a set of elements into two sets, affected and unaffected. It comparse multiple evidences to determine which elements of a network are most likely affected by a perturbation of a system. The input to this program is a file containing a matrix of significances for observations for evidence types (columns) and network elements (rows), in a specific format. Missing data is allowed; a missing significance value is denoted by the value "-1", a "null" string, or (in certain cases) an empty cell in the input file. The smaller a significance value for an observation, the **more** likely it is that the associated element is affected by the perturbation of the system; in this sense, the significance is analogous to a probability that a given observation would occur, given that the associated element is *not* a member of the set of affected elements. The significances may be calculated using the Significance Calculator, or they may be generated by any other procedure that can assign a statistical likelihood or probability. Each evidence type is assigned a weight, based on consistency with the other types of evidence. The weights are used to compute an **effective significance** for each significance value in the matrix.

The following list describes the various components and controls of the Evidence-Weighted Observer program window:

Load sigs
Loads the file of significances into the Input Data Table.

Delimiter
This drop-down list is used to specify how your input data file (and output data file) are delimited. The three choices are "tab", "space", and "comma". Comma-delimited files are recommended for reasons of unambiguity and compatibility with spreadsheet programs. If you choose "space", be advised that there is a restriction on how you can specify an empty cell (missing data). With a "space" delimiter, an empty cell must be denoted by the string "null" (without double quotes). Note that the string is case-sensitive, so you cannot use the string "Null" or "NULL". Once you have loaded your input data file, you may change the delimiter that you use to save the output file, if you wish.

Input data table
This table displays the significances that have been loaded into the program, in the top half of the program window. Cells for which there is no significance (denoted by a "-1", "null", or an empty cell in the input data file) are shown as empty in the table. The column names are evidence types, and the first row contains the network element names. All other cells are either numeric values denoting the significance of the evidence and element pair.

Number of bins for significance distributions
This type-in box is used to specify how many bins are to be used for the calculating the nonparametric distribution of overall significance values of the elements. The program suggests a default value; you may change it to any integer greater than 1.0.

Combined significances quantile cutoff
This type-in box is used to specify a threshold for overall (combined) significances. Any element whose overall significance quantile is less than this threshold, will be placed to in the set of affected elements, provided the significance value is less than the "critical significance". The critical significance is the significance value at which the distributions of affected and unaffected element signficances overlap and are equal. This cutoff parameter should be kept very small. If you are not sure what value to use, just use the default.

Area separation threshold (from unity)
This type-in box is used to specify the criterion for exiting the iterative algorithm. The algorithm uses an objective function, which is defined as the fraction of the area of the distribution of significances of affected elements which does not overlap the distribution of unaffected elements (objective function). Then this objective function changes by an absolute amount smaller than the "area separation threshold" from one iteration to the next, the algorithm will exit. This value should be kept small, and is required to be positive-definite. If you are not sure what value to use, just leave the default.

Effective evidence weighting scheme

This drop-down list is used to specify how the effective significance is to be calculated from the raw (unweighted) significance. The three choices are "linear", "power", and "uniform". The "linear" choice means that the effective significance `Seff` is calculated as:

```
Seff = b + W*S,
```

where `b` is the bias term, `W` is the weight, and `S` is the unweighted significance for the element and evidence. The weights are normalized, i.e., the sum of the weights over all evidences is equal to 1.0. The bias of an evidence is 1.0 minus the weight, divided by the sum of the same quantity over all evidence types. The "power" choice means that the effective significance `Seff` is calculated as:

```
Seff = S^W
```

where the "`^`" character denotes exponentiation.

The "uniform" weight type means that each evidence will be given equal weight. This can be used in certain cases where Pointillist might have difficulty ascertaining the statistical weights from the data distributions.

The default evidence weight type is "power".

Initial quantile threshold for affected

This type-in box is used to specify the quantile for the significance cutoff used to build the initial set of "putatively affected" elements. An element is placed in the initial set of putatively affected elements if there is at least one evidence for which the quantile of the element's significance is less than this "initial quantile threshold for affected"; thus, the initial set is built greedily, and will likely be much larger than the true set of affected elements. This value is initially set to a small default value; you may change it to any positive-definite value less than 1.0. If you are unsure what value to use, just use the default.

Fraction of elements to move in each iteration

This type-in box is used to specify the initial fraction of elements that can be changed from "affected" to "unaffected" within a given iteration. It controls how aggressively the algorithm will prune the putative set of affected elements, within a given iteration. As the algorithm proceeds, the fraction of elements that will be pruned from the affected set within a given iteration decreases; this parameter controls the starting fraction to remove. The default is a small fraction, but you may change it to any value between 0.0 and 1.0 (exclusive). If you are unsure what value to use, just use the default.

Smoothing length for significance distribution

This type-in box is used to specify the smoothing length for obtaining the smoothed, normalized nonparametric distribution of overall significances of elements. Specifically, this parameter is used as the standard deviation of the Gaussian kernel density function for smoothing the nonparametric distribution of overall significances. The default is a small fraction, but you may set to any value greater than 0.0. If you are not sure what value to use, just use the default.

Maximum number of iterations

This type-in box is used to specify the maximum number of iterations for the inference algorithm. If this box is left empty, there will be no limit on the number of iterations of the inference algorithm. In that case, the algorithm will complete only when the change in the "objective function" is less than the threshold specified in "area separation threshold". If a maximum number of iterations is specified using this field, it must be a positive integer.

Infer affected elements

This button starts the inference algorithm. Depending on the size of the input data table, and the parameters you chose for "fraction of elements to remove in each iteration" and "initial p-value cutoff for a single evidence type", the algorithm may take a long time to complete. If the algorithm fails to complete in a timely fashion, the best way to stop it is to close the program's window in your computer's window manager. When the algorithm completes, the results are displayed in the Results

Data Tables and the Results Statistics frames.

Results tables

The results of the inference algorithm are displayed in three tables in the right side of the bottom half of the program window.

- The first table has the columns "evidence name" and "weight". It indicates the weight used for the final iteration of the inference algorithm, for each evidence type. A low weight indicates less reliable data, as determined by consistency with the other evidence types.
- The second table (below the weights table) is the "iteration summary table". This table contains an iteration-by-iteration description of the number of putatively affected elements, and the number of "probable false negative" elements. Probable false negative elements are members of the set of "putatively unaffected" elements, that nonetheless have high significance. It should be noted that probable false negatives are not transferred to the "putative affected set" until all iterations have completed. Thus, the "probable false negative" column is a running estimate of the number of false negatives. The number appearing for the last iteration indicates the number of elements actually moved from the unaffected to the affected set, because they were identified as probable false negatives.
- The third table has the column names "element name", "affected", and "overall significance". The "affected" column is a boolean value indicating whether the element is in the final putative affected set. The "overall significance" column gives the final overall significance value for the element; a smaller significance value indicates greater likelihood that the element is in the true affected set. You may sort the results in the rightmost table, by clicking on the table header. A control-click in the table header allows for sorting based on multiple columns. Repeatedly clicking (or control-clicking) will cycle through the three different sorting states: none, ascending, and descending.

Results statistics

This set of fields shows some statistics pertaining to the inference algorithm.

- **num iterations**
  The "num iterations" field indicates how many iterations of the inference algorithim occurred before it exited.
- **final separation**
  The "final separation" field indicates the fraction of the area of the (smoothed, nonparametric) overall significance distribution for the putative affected elements that is nonoverlapping with the distribution for the unaffected elements. This value should be very close to 1.0.
- **alpha parameter**
  The "alpha parameter" field gives an indication of data independence between the different evidence types, based on a ratio of the average of the product (over the evidences) of significances (over the putatively unaffected elements) to the product (over evidences) of the average (over unaffected elements) of the significances. It should be very close to 1.0; if it is not within the range 0.9-1.1, the value will be displayed in red, to indicate a warning to the user.
- **num affected**
  The "num affected" field indicates how many elements were in the final putative set of affected elements.

Save weights

This button saves the table of evidence-specific weights to a file. The currently selected delimiter (see above) is used to delimit the columns of the data file that is produced. The first row contains the column header names. The first column contains the evidence names. The second column contains the weights. If an evidence name contains a delimiter character, the delimiter characters appearing in the element name will be converted to underscore ("_") characters.

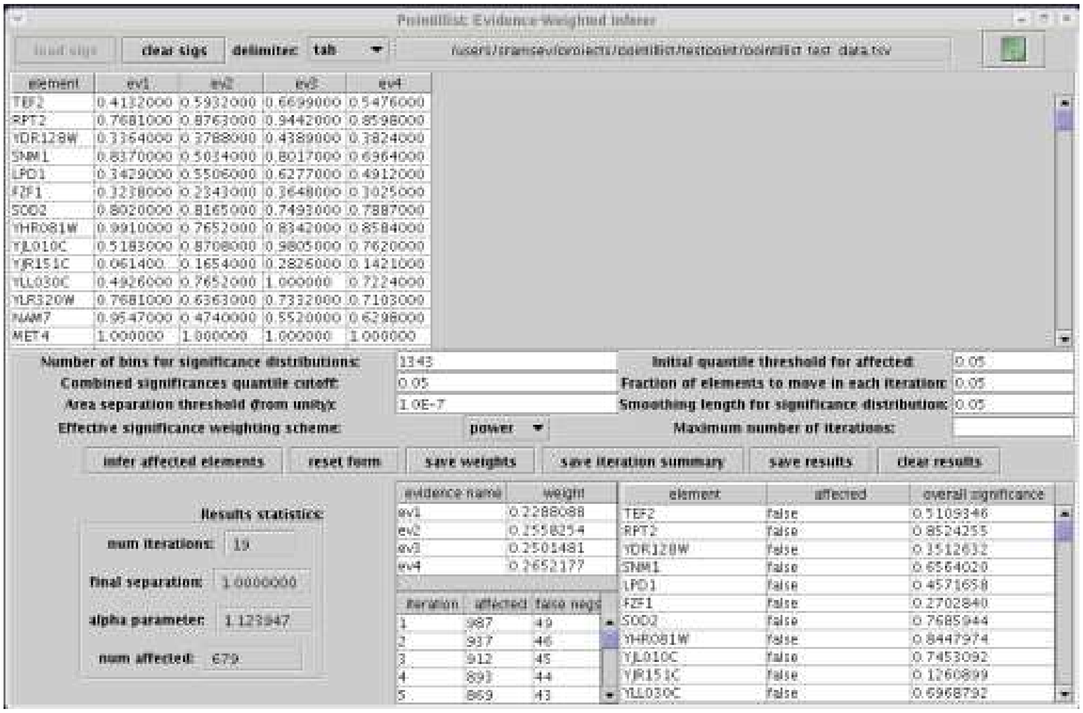
Save iteration summary

This button saves the table containing the iteration summary, to a file. The currently selected delimiter (see above) is used to delimit the columns of the data file that is produced. The first row contains the column header names. The first column contains the iteration number (starting from 1). The second column contains the number of putatively affected elements, after that iteration. The third column contains the number of probable "false negative" elements after that iteration.

Save results

This button saves the results table (the table containing the list of elements, their boolean membership in the putative affected set, and their overall significances) to a file. The currently selected file delimiter (see above) is used to delimit the columns of the data file that is produced. The first row of the data file contains the column header names. The data is saved in the same sorting order as it is displayed in the results table. If you change the delimiter type between when you load the data file and when you save it, and if an element name contains the new delimiter character, the delimiter characters appearing in the element name will be converted to underscore ("_") characters.

Here is a screen shot of the Evidence-Weighted Inferer program:



delimiter: tab

/users/sramsey/projects/pointillist/testpoint/pointillist_test_data.tsv

element	ev1	ev2	ev3	ev4
TEF2	0.4132000	0.5932000	0.6699000	0.5476000
RPT2	0.7681000	0.8763000	0.9442000	0.8598000
YDR128W	0.3364000	0.3788000	0.4389000	0.3824000
SNM1	0.8370000	0.5034000	0.8017000	0.6964000
LPD1	0.3429000	0.5506000	0.6277000	0.4912000
FZF1	0.3238000	0.2343000	0.3648000	0.3025000
SOD2	0.8020000	0.8165000	0.7493000	0.7887000
YHR081W	0.9910000	0.7652000	0.8342000	0.8584000
YJL010C	0.5183000	0.8708000	0.9805000	0.7620000
YJR151C	0.061400	0.1654000	0.2826000	0.1421000
YLL030C	0.4926000	0.7652000	1.000000	0.7224000
YLR320W	0.7681000	0.6363000	0.7332000	0.7103000
NAM7	0.9547000	0.4740000	0.5520000	0.6298000
MET4	1.000000	1.000000	1.000000	1.000000

Number of bins for significance distributions: 1343

Combined significances quantile cutoff: 0.05

Area separation threshold (from unity): 1.0E-7

Effective significance weighting scheme: power

Initial quantile threshold for affected: 0.05

Fraction of elements to move in each iteration: 0.05

Smoothing length for significance distribution: 0.05

Maximum number of iterations:

Results statistics:

num iterations: 19

final separation: 1.0000000

alpha parameter: 1.123947

num affected: 679

evidence name	weight
ev1	0.2288088
ev2	0.2558254
ev3	0.2501481
ev4	0.2652177

iteration	affected	false negs
1	987	49
2	937	46
3	912	45
4	893	44
5	869	43

element	affected	overall significance
TEF2	false	0.5109346
RPT2	false	0.8524255
YDR128W	false	0.3512632
SNM1	false	0.6564020
LPD1	false	0.4571658
FZF1	false	0.2702840
SOD2	false	0.7685944
YHR081W	false	0.8447974
YJL010C	false	0.7453092
YJR151C	false	0.1260899
YLL030C	false	0.6968792

Matlab Codes

The Pointillist algorithms are also available as a set of Matlab M-files. These files can be found in the `matlab` subdirectory of the directory where you installed Pointillist. Note that these Matlab codes require the Matlab Statistics Toolbox.

```
    basisexp.m: select a dataset to be used for p-value normalization

    pscalef.m: normalization of p-values (correct non-ideality of p-value distributions or correlation issues if any).

               When applying "pv2.m" to the p-values obtained from real high-throughput datasets,
               we recommend to run the two functions above to normalize p-values before running "pv2.m"

    pv2.m: a wrapper to call esa.m and genwfunf.m depending upon the methods mentioned in the main text of the paper

    esa.m: enhanced simulated annealing that searches for the optimal weight and parameters.

    genwfunf.m: select elements given a weight vector and alpha

    mcmc.m: Monte Carlo simulation to generate random numbers for each integration method (for example, Fisher’s method)

    nprampv2.m: non-parametric method for integrating datasets

    paretospace.m: perform a multi-objective optimization on the Pareto space

    supsmu.m: non-parametric smoothing algorithm.

    nwpv2.m: non-weighted integration methods.
```

These Matlab codes have been tested with Matlab R13 and R14. They are not used by the Java implementation of Pointillist, but are a separate implementation of the Pointillist algorithms.

Getting Help

If you find that the Pointillist program does not function in accordance with the descriptions in this manual, or if there are sections of this manual that are incorrect or unclear, the author would like to hear about it, so that we can make improvements and fix bugs in the software. Furthermore, the author would appreciate feedback regarding new features or improvements that would be useful to users of this software. Before e-mailing the author, it is a good idea to check the Pointillist application home page to see if a new version has been released, in which your specific problem may have been corrected. All releases are documented in the "version history" page accessible from the home page. The best way to contact the author is to send e-mail to:

`pointillist at systemsbiology.org`.

The author will attempt to respond to your e-mail as quickly as possible.

If you are reporting a bug, or something that you suspect is a bug, please provide as much information as you can about your specific installation of the Pointillist program. In particular, please provide us with the version number of the Pointillist program that you are using, the type and version number of the Java Runtime Environment that you are using (e.g., Sun JRE version 1.4.1), and your operating system type and verion (e.g., Red Hat Linux 8.0). Furthermore, if the problem is with a specific model definition file, please send us the model definition file, and any model definition files that it includes. If the problem that you encountered generated a "stack backtrace" on the console, please include the full stack backtrace text in your bug report. Please also send us the text of any error message that you may have been reported by the application, in a dialog box or the like. Providing this information will dramatically increase the likelihood that the author will be able to quickly and successfully resolve the problem that you are encountering.

Last updated: 2006/01/30 20:19:27

Please e-mail comments or corrections regarding this document to: `pointillist at systemsbiology.org`