PFunc

PFunc is a set of tools for fitting and analyzing function-valued traits. It was designed for quantifying preference functions (hence the name) in biological studies of sexual selection and evolution of mate preferences. PFunc accepts a bivariate dataset composed of stimuli and responses to those stimuli as its input. It then fits cubic splines to these data using the gam function in R, and it displays the curves, along with several useful measurements of the shape of the curve. Users have several options for adjusting the curves and the tools used to measure the curves, and for outputting results for further analysis.

This file was last updated 2017-05-18 for PFunc version 0.10.0.

Table of Contents

Files

`PFunc.py` - the main file to run for the full PFunc GUI experience

`PFunc_RCode.R` - the supporting R code that fits the splines and extracts the useful metrics. This code *can* be run on its own in R without the GUI

`README.pdf` - important information for installing and using the program

`demo_data_horizontal.csv` - example data in one of two possible layouts

`demo_data_vertical.csv` - example data in one of two possible layouts

`PFuncIcon.png` - icon file

`PFuncPath.txt` - may help users who encounter path-related errors

`COPYING.txt` - the full GPLv3 license

Setup

In the fullness of time, we aim to wrap PFunc up in a single easy-to-use executable file. For now however, some hands-on setup is required. Follow these steps to setup PFunc on Windows, Mac OS X or Linux systems.

1. Install and Set Up R

1. Visit https://cran.r-project.org/ and follow the links to download the latest version of R for your operating system. Install R as you would a normal program.
   Note for Windows users: The setup wizard will give you the option of installing the 32-bit files or the 64-bit files. Install both.
   Note for Linux users: You may instead install R from the command line.

2. Install the mgcv package in R by opening R and entering the following command: `install.packages("mgcv")` . You will be prompted to select a mirror from which to download the package. Select an option that is relatively close to your location. Follow any instructions R gives you for installing the package.

2. Install and Set Up Python

1. Visit https://www.python.org/ and follow the links to download the latest version of Python 3.x for your operating system. Install python as you would a

normal program.
Note for Windows users: the first screen of the installation wizard will ask if you want to add Python to PATH. Make sure to select this option.
Note for Linux users: you may instead install Python from the command line.

2. Install the matplotlib Python library. To do this, open up the terminal (if you are not sure how, run a search for "terminal" or "command prompt" on your computer), and run one of the following commands.
For Windows users: enter this command: `pip install matplotlib`. If that doesn't work, try calling pip this way instead: `py -m pip install matplotlib`
For Mac and Linux users: enter this command: `pip3 install matplotlib`

3. Install the rpy2 Python library.
For Windows users: Go to http://www.lfd.uci.edu/~gohlke/pythonlibs/#rpy2 and download one of the rpy2.whl files. More likely than not, you're going to have to try a few before you find one that works, so be patient here. These files are named in the following format: `rpy2-[chunk1]-[chunk2]-[chunk3]-[chunk4].whl`. [chunk1] is the version of rpy2--generally the more recent the version, the better. [chunk2] points to the version of Python that this file works with, for example, cp35 means it works with Python 3.5; cp36 works with Python 3.6, etc. Choose one that matches your version of Python. [chunk4] tells you something about the processor architecture that the file is expecting to run on. `win32` means it's looking for a 32-bit processor, and `win_amd64` means it's looking for a 64-bit processor. Note, however, that depending on how Python or R are set up on your system, you may have to use a file marked as `win32` even if you have 64-bit Windows.

   - Download the rpy2.whl file, and take note of the directory path where you save it.
   - Enter the following command in the command prompt (like in the previous step), substituting the directory path below for the one where you saved the rpy2.whl file:
     `pip install "C:\Users\name\Downloads\rpy2-2.8.3-cp35-cp35m-win32.whl"` (and again, if you can't call pip directly, try `py -m pip install...` instead).
   - If you encounter an error, it's likely you'll have to download a different rpy2.whl file from the list and try installing it using these same steps.

   For Linux and Mac users: The process is much simpler. Enter the following command in the terminal, just like with the previous step: `pip3 install rpy2`

A note on package versions

PFunc was written with particular versions of software and packages in mind. Given that PFunc depends on these packages to run, it is possible that at some point in the future, these packages could be updated in a way that negatively impacts the functionality of PFunc. If this happens, you have two options:

- Check http://github.com/joccalor/pfunc for updated PFunc code that is compatible with the new versions of the packages.

or:

- Search online for old versions of the packages that work with this version of PFunc. Below is a list of versions known to work with PFunc. These packages should all be archived online. Don't worry if you cannot find these exact versions; other close versions will probably function just fine.

  - R 3.3.1
  - mgcv 1.8-17
  - Python 3.5.2
  - matplotlib 1.5.3 and 2.0.1
  - rpy2 2.8.5

  For packages installed through pip, you can specify a version number with the double equals sign (==). So for example, if you were to want to uninstall your current version of matplotlib and install version 1.5.3, you would use these commands:
  ```
  pip uninstall matplotlib
  pip install matplotlib==1.5.3
  ```

Usage

Below is a bare-bones overview of using PFunc. For more details on usage as well as general theory underlying the study of preference functions, check out our paper "Describing mate preferences and other function-valued traits" by Kilmer, Fowler-Finn, Gray, Höbel, Rebar, Reichert & Rodríguez in the Journal of Evolutionary Biology.

Data Input

PFunc is expecting data with two main components: a set of **stimuli** (plotted along the *x*-axis) and a set of **responses** to those stimulus values (plotted on the *y*-axis).

PFunc needs a bare minimum of three data points to fit one spline (although we HIGHLY recommend using more than three data points). This means that if you are splining at the individual level (which is what we tend to do), each individual must be tested against a range of stimulus values.

The other main way to create preference functions is to test a sample of individuals across a range of stimuli, and then use the combination of those

individual responses to create one spline. This would be the way to go in a species where you could only test each individual once. If you use this second option, then be aware that you are no longer splining at the individual level, but at a group level, and so it will be necessary to communicate this to PFunc by grouping individuals in your data file. If you are using the horizontal data file layout (described below), then you can create one response column with each of the different responses. If you are using the vertical data layout (described below), then you can create a new column to code for group identity. When you open the vertical file in PFunc, then you can tell PFunc to spline according to group identity, rather than individual identity.

You have the option of formatting your data in one of two ways:

- In the **horizontal** layout, the first column of data contains the stimulus values, and each subsequent column contains the responses of an individual to those stimulus values. See the `demo_data_horizontal.csv` file for an example.
- In the **vertical** layout, data are organized into a minimum of three columns: one with individual identities, one with stimulus values presented to those individuals, and one with the responses of those individuals to those stimuli. See the `demo_data_vertical.csv` file for an example.

Regardless of which layout option you choose, you must save your data as a **.csv** file (any type of .csv will work).

Running the PFunc GUI

Startup

You have several different options for running PFunc. Make sure that `PFunc.py` and `PFunc_RCode.R` are together in the same directory. If you encounter an error, see the Troubleshooting section below.

- **Option 1: Run PFunc through IDLE**
  The most straightforward way to run PFunc is through IDLE. Right-click (or the equivalent on a Mac) on the `PFunc.py` file, and choose to "Open With..." IDLE (or "Edit With Idle" on Windows if that option is available). This will start up IDLE with the code for PFunc displayed in a window. Click on the window of colorful code to make sure it is the current active window, and then run the code either by pressing F5 or by clicking Run > Run Module in the menu at the top bar.
- **Option 2: Run PFunc from the command line**
  - Take note of the directory where PFunc is saved.
  - Open up an instance of your terminal (command prompt in Windows) and navigate to the PFunc directory using the `cd` command. Examples: `cd "C:\Users\name\PFunc"` in Windows; `cd "/Users/name/PFunc"` in Mac OS X; `cd "/home/name/PFunc"` in Linux.
  - Then enter this command for Windows: `python PFunc.py` (or if that doesn't work, try: `py PFunc.py`), or this command for Mac or Linux: `python3 PFunc.py` .

- **Troubleshooting**
  You might encounter an error while trying to run PFunc if rpy2 has trouble finding R on your computer. One sign of this error is if the error message mentions R_HOME or PATH. To solve this, you'll have to find where R is installed on your computer and follow the instructions in the `PFuncPath.txt` file that accompanies this program.

  As stated on the R downloads page, Mac users may encounter errors with R if XQuartz is not installed. In PFunc, problems are most likely to be encountered when trying to output graph files without XQuartz installed. XQuartz can be downloaded from https://www.xquartz.org/.

The Interface

Once you open your data file, PFunc will display a page of graphs, each one showing data from a single individual with a spline fit through the points. A page displays up to nine individuals, and you can navigate to different pages with the controls at the bottom of the screen.

You can select a graph by clicking on it, and you can enlarge a graph by double-clicking on it.

When a graph is selected, its smoothing parameter is displayed in the Smoothing box on the control panel on the right-hand side of the window, and several important metrics are displayed in the Summary box right below the Smoothing box.

Smoothing Parameter

The smoothing parameter controls how stiff or how wobbly a spline is. A low smoothing parameter yields a wobbly spline, and at extremely low values, the spline runs through every data point. A high smoothing parameter yields a spline that is a gentle curve, and at extremely high values, the spline is a straight line. The gam function in R (which fits the splines) chooses an optimal smoothing parameter for you. In most cases, it is not necessary to alter the smoothing parameter. However, should you wish to change it, you have two main options:

- **Specify your own smoothing parameter.** You can use the "-" and "+" buttons next to the smoothing parameter to increase and decrease the value. You may also type in your own smoothing parameter and set it by pressing the Enter key. When a smoothing parameter has been altered, the small magenta dots in the bottom left corner of the graphs change to cyan. If at any point you want to revert to the default smoothing parameter, simply press the Reset button in the Smoothing box (this will also change the cyan dot back to magenta).
- **Adjust the limits imposed upon smoothing parameters.** PFunc was built with the philosophy that preferences function splines should be neither too wobbly nor too stiff. Therefore, by default, PFunc restricts all smoothing parameter values to be between 0.05 and 5. We found this to be a reasonable range for the types of data we were working with, but you may find a different range to be more suitable for your data. You can change this range in the Smoothing Limits setting (described below). Alternatively, you may decide that you want no limits on your smoothing parameters, and

instead want to see what default the program gives you, no matter how big or how small. You have the option of turning off smoothing parameter limits altogether (also described in the Smoothing Limits setting below). When you make changes to the Smoothing Limits setting, only the graphs with magenta (and not cyan) dots will change. This prevents you from accidentally wiping away any intentional changes to specific smoothing parameters.

Summary

The summary window displays several useful measurements of the spline in the selected graph:

- **Peak Preference:** This is the *x*-axis value that corresponds to the highest point of the curve, as illustrated by the red vertical lines on the graphs.
- **Peak Height:** This is the *y*-axis value that corresponds to the highest point of the curve.
- **Tolerance:** This is the width of the curve at a certain height. By default it is measured 1/3 of the way down from the peak of the curve to zero. See the Tolerance settings below to change the proportion of the drop or to measure tolerance at a set *y*-axis value for all curves.
- **Strength:** This is a measure of how far the curve drops away from the peak. By default, the measure of strength depends on the elevation of the spline, because the same absolute drop in a curve will be a larger proportion of a curve that is close to the baseline, compared to a curve that is high up. This height-dependent strength is calculated as the squared coefficient of variation of the *y*-axis values along the spline. A height-independent option may be used instead, which calculates strength as the standard deviation of the *y*-axis values along the spline divided by their range.
- **Responsiveness:** This is the average height of the curve.
- **Smoothing:** This displays the smoothing parameter used to generate the curve with the above values. Most of the time it is the same as the value in the Smoothing box.

Settings

You have access to a variety of settings that give you control over how your data are analyzed. Below are descriptions of the available settings. PFunc will always start up with the default settings, but you can save any changes you make to the settings through the File menu ("Save Current Settings"), and then load them again later ("Load Previous Settings"). You can also revert to the default settings ("Restore Default Settings").

- **View** Here you can toggle the visibility of five different things on or off, depending on how you want to view your graphs. If you choose to export your spline graphs, these View settings will carry over to the output file.
- **Smoothing Limits** Here you can adjust the limits to the smoothing parameters that are chosen for your splines. By default, PFunc limits them to values between 0.05 and 5, but you have full control over the minimum and maximum values here, and you also have the option to turn these limits off entirely with the checkbox.
- **Find Local Peak** Here you can narrow the range in which PFunc searches for peaks. If it finds a peak in your specified range, it uses that value. If no peak is found, then PFunc uses the regular peak instead.
- **Tolerance** This gives you fine control over how tolerance (the width of the curve at a given height) is measured.
  - First you have the option of measuring tolerance relative to the height of the peak (e.g. 1/3 of the way down from the top), or measuring it at a set *y*-axis value. To switch between these two options, click one of the two radio buttons next to the appropriate settings: the radio button next to "Drop from peak" is for the relative option, and the radio button next to "At set value" is for the absolute option.
  - Once you've selected *how* tolerance should be measured, you can control *where* it is measured. If you've chosen the relative option, you can control the drop from the peak. This takes the distance from the peak of the curve to the floor (which you can also set), and measures tolerance at a particular proportion of that distance down. If, on the other hand, you've chosen to measure tolerance at an absolute height, you then can specify that exact height with the "At set value" setting.
  - Finally, you can decide whether you want a broad measure of tolerance or a strict one. The difference comes into play when there are secondary peaks in your curves. If you choose the broad option, then tolerance will be measured as the total width of the curve under the primary peak plus any relevant secondary peaks. If you choose the strict option, then tolerance will be measured as the width of the curve *only* under the primary peak.
- **Strength** Here you have the option of using the height-dependent measure of strength (recommended) or the height-independent measure of strength (see above for brief description of the two, and see our paper for a more complete discussion).

Group-Level Splines

You can combine individual splines to form group-level splines. This can come in handy if you want to generate splines at the replicate-, treatment-, family-, population-, or species-level. Group splines can even be combined to form higher-order group splines.

To do this, go to Advanced > Construct Group-Level Spline... . The pop-up window allows you to name your new group and select which individuals belong in the group (click and drag or use the Shift and Ctrl keys to select multiple individuals). Once you are finished, press Okay, and your new group-level spline will be added alongside your other splines.

Note that this does not affect your input data file; if you want to retain these values, you'll need to output them (see below). Also note that group-level splines may be best fit with lower smoothing parameters than individual-level splines.

Message Log

PFunc keeps track of all its warnings and confirmations, even ones that it doesn't explicitly make pop-ups for. To see the running log of messages, go to Advanced > Show Message Log.

Output

Here are descriptions of the various output options available under the File menu.

- Output Spline Figures: Creates a .pdf, .eps or .svg file with all the graphs using the current smoothing parameters.
- Output Spline Summaries: Creates a spreadsheet that contains all of the information in the Summary window for every individual.
- Output Spline Points: PFunc extracts the *y*-values at 201 evenly-spaced points along the curve, and it saves these as a spreadsheet. This is useful if you want to plot your curves in a different program.
- Output Tolerance Points: Creates a spreadsheet containing all of the *x*-axis values that correspond to the upper and lower limits of tolerance--that is, the start and stop points of the horizontal blue lines in the graphs.

Running PFunc from the R Command Line

If you are comfortable working in the R command line environment, you may use Pfunc without the GUI. Note that when PFunc is used this way, data **must** be set up in the horizontal format (as in `demo_data_horizontal.csv`), never in the vertical format (as in `demo_data_vertical.csv`).

Startup

Open R. Set the working directory to wherever the `PFunc_RCode.R` file is saved on your computer.

- One way to do this is by finding the directory path and entering the command `setwd(directory/path)` . You can check your current working directory with the command `getwd()` .
- There is another way to do this if you are not running R directly from your terminal. Depending on your version of R, you will find the option either under "File > Change dir..." or "Misc > Change Working Directory..."
- Load the script into R with this command `source(PFunc_RCode.R)`
- Load your data into R with a command like `mydata <- read.csv("datafile.csv")`
- Now you can run the default analysis on your data with the command `PFunc(mydata)` . See below for a full set of options.

Arguments of PFunc

Here is a list of arguments that you can specify when calling PFunc directly from R.

- `input.data` - the name of your dataset in the R environment (no default)

- `diagnose.col` - an optional argument used only when assessing individual functions. To see the graph and metrics for a single individual without creating a new file, set this number equal to the corresponding column in your datasheet (default = 0).

- `diagnose.sp` - an optional argument used only when assessing individual functions. By setting this to a positive number, you can specify the smoothing parameter of individual functions that you are examining (default = -1).

- `auto.sort` - an option to sort the data by increasing values of the stimulus variable (default = TRUE).

- `blacklist` - if there are columns of your data that you wish to exclude from analysis, place their numbers in this argument with the concatenate `c()` function. For example, to exclude the third and fourth columns, set `blacklist = c(3, 4)` (default = NULL).

- `whitelist` - if there are only a few specific columns of your data that you wish to analyze and exclude the rest, you can place the included columns' numbers in the whitelist with the concatenate `c()` function. For example, to include only the third and fourth columns in analysis, set `whitelist = c(3, 4)` (default = NULL).

- `sp.binding` - when true, it constrains the auto-generated smoothing parameters to be between `min.sp` and `max.sp` (see below) (default = TRUE).

- `sp.assign` - allows you to specify your own smoothing parameters (sp) for individual columns rather than letting the script generate optimized values. Must be in the form of a two-column matrix in which the first column contains the column numbers for which you want to specify smoothing parameters. The second column contains the corresponding desired smoothing parameters. For example, `new.sp <- matrix(c(2, 1, 3, .5, 4, 0), ncol=2, byrow=T)` . Here, the individual in input.data column 2 gets an sp of 1, the individual in input.data column 3 gets an sp of .5, and the one in column 4 gets an sp of 0. To run with these values, set `sp.assign = new.sp` (default = 0).

- `max.sp` - if you are constraining auto-generated smoothing parameters with the `sp.binding` argument (above), `max.sp` is the upper limit (default = 5).

- `min.sp` - if you are constraining auto-generated smoothing parameters with the `sp.binding` argument (above), `min.sp` is the lower limit (default = 0.05).

- `peak.within` - tells PFunc where to search for local peaks. If a single number, x (0 < x <= 1), is passed for this argument, PFunc will search in the central x proportion of the stimulus range. For example, if this argument is passed with 0.7, then PFunc will look in the central 70% of the function for a peak. If it finds one, it uses it. If it doesn't find one, it uses the regular peak. If two numbers are passed instead (e.g., `peak.within = c(120, 150)`), then those will serve as the lower and upper bounds between which PFunc will look for local peaks.

- `summary.out` - filename for the output summary that includes peak, tolerance, and other measures of each function. If FALSE, no file will be made (default = "spline_summaries.csv").

- `graph.out` - filename for the output graphs for all analyzed functions. If FALSE, no file will be made (default = "spline_graphs.pdf").

- `points.out` - filename for the output points that make up each function. Useful for plotting functions in external programs. If FALSE, no file will be made. You can specify the number of points you want to output for each function in the `predictions` argument below (default = FALSE).

- `max.y` - for the output graphs in `graph.out` (above), this argument specifies the height of the vertical axis. Leave at default value of 0 for automatic.

- `pdf.row` - for the output graphs in `graph.out` (above), this argument specifies the number of rows of graphs each page of the pdf will have (default = 4).

- `pdf.col` - for the output graphs in `graph.out` (above), this argument specifies the number of columns of graphs each page of the pdf will have (default = 3).

- `graph.points` - an option to turn data points on or off in the output graphs (default = TRUE).

- `graph.peak` - an option to turn on or off a red vertical line at the peak of the function (default = TRUE).

- `graph.tol` - an option to turn on or off a blue horizontal line showing the width calculated for tolerance (default = TRUE).

- `graph.sp` - an option to print the smoothing parameter of each function above its corresponding graph (default = FALSE).

- `graph.se` - an option to plot the standard error around the spline (default = FALSE).

- `drop` - specify how far down from the peak of the function that tolerance (the width) should be measured (default = 1/3).

- `tol.mode` - how would you like the script to calculate tolerance? The default value "broad" (alternatively "norm") looks at the width of the function under the main peak plus any other areas at that same *y*-value that may be under secondary peaks. The alternative value "strict" only counts the width of the part of the function that contains the main peak and ignores other peaks, even if they rise above the chosen *y*-value.

- `tol.floor` - this specifies the baseline of the data. It is used in conjunction with `drop` to determine the height where tolerance should be measured. For example, if `drop` equals 1/3, and `floor` equals 0, then tolerance will be measured at 1/3 of the distance from the peak to 0 (default = 0).

- `predictions` - if you want to extract points along your splines using `points.out`, this argument tells the script how many points you would like. The default value 0 outputs points only at the input stimulus values. Anything greater than 0 will output that many points evenly spaced along the *x*-axis.

- `ghost` - if you are specifying your own smoothing parameters with the `assign.sp` argument, and if you are creating an output graph file with the `graph.out` argument, then setting ghost to TRUE will plot splines with the original default smoothing parameters in gray alongside your new splines. This is useful for comparing the effects of your altered smoothing parameters (default = FALSE).

- `allfromsplines` - an option to specify how you would like strength and responsiveness to be calculated. With the default value of TRUE, they will be calculated from the splines. Change this to FALSE only if you want to calculate strength and responsiveness from your input data points.

Examples

The following examples assume that your data file is called "mydata" in the R environment.

- To output spline summaries and a pdf of the graphs using automatic smoothing parameters bound between 0.05 and 5, use this command `PFunc(mydata)`

- To look at an individual spline (say, the one in the third column of your spreadsheet) without outputting any files, use this command `PFunc(mydata, 3)`

- To see what that individual spline would look like with a new smoothing parameter (say, 0.7), use this command `PFunc(mydata, 3, 0.7)`

- To keep this change for the final output, start by constructing a matrix as described in `sp.assign` above. Then set `sp.assign` to the name of the matrix, like this:

```
new.sp <- matrix(c(3, 0.7), ncol=2, byrow=T)
PFunc(mydata, sp.assign=new.sp)
```

- By default, the height at which tolerance is measured is a proportion of the distance between the peak of a curve and the value of the `floor` argument. If you prefer to measure tolerance at the same *y*-axis height across all individuals (say, 2.5), regardless of the height of the peak, set `floor` equal to your desired height, and set `drop` equal to 1, like this `PFunc(mydata, floor=2.5, drop=1)`.

Acknowledgements

Contact

For comments or questions, contact Joey Kilmer at joeykilmer@gmail.com.
Get the latest version of PFunc at https://github.com/Joccalor/PFunc/releases/latest

License