# DemoGrapher Tutorial

Ethan Jewett
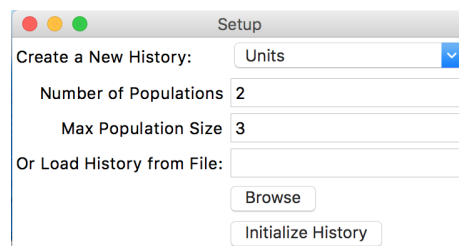
2/4/2017

## 1 Before you begin

This tutorial assumes that you already have DemoGrapher installed. See the `README.rst` file for instructions on downloading and installing. Once you have a Python 2.7 distribution on your computer, installation should be as simple as typing `conda install -c jewett demographer` or `pip install demographer_popgen`.

## 2 Starting DemoGrapher

To start DemoGrapher, type `demographer` at a terminal prompt[1]. This will give you a startup screen that looks like this:



This allows you to set up a generic population history that you'll modify later[2]. The current window says that we want a demographic history with 2 populations, each with a maximum size of 3-times some reference population size[3].

---

[1]if you're running DemoGrapher in a virtual environment, you'll first need to do `conda activate your_environment_name` or `source path_to_virtual_environment/activate`

[2]If you have saved a population history from a previous session, you can browse for it using the `Browse` button or enter the file path in the `Or Load History from File` entry field.

[3]You can change the units to something more intuitive using the `Units` menu. For example, if you want to work with absolute effective population sizes and years instead of coalescent units, you can choose `Eff. Size and Years` and text fields will open up for you to enter effective sizes and years per generation.

To start a DemoGrapher session, click `Initialize History`. You'll get a screen that looks like this:
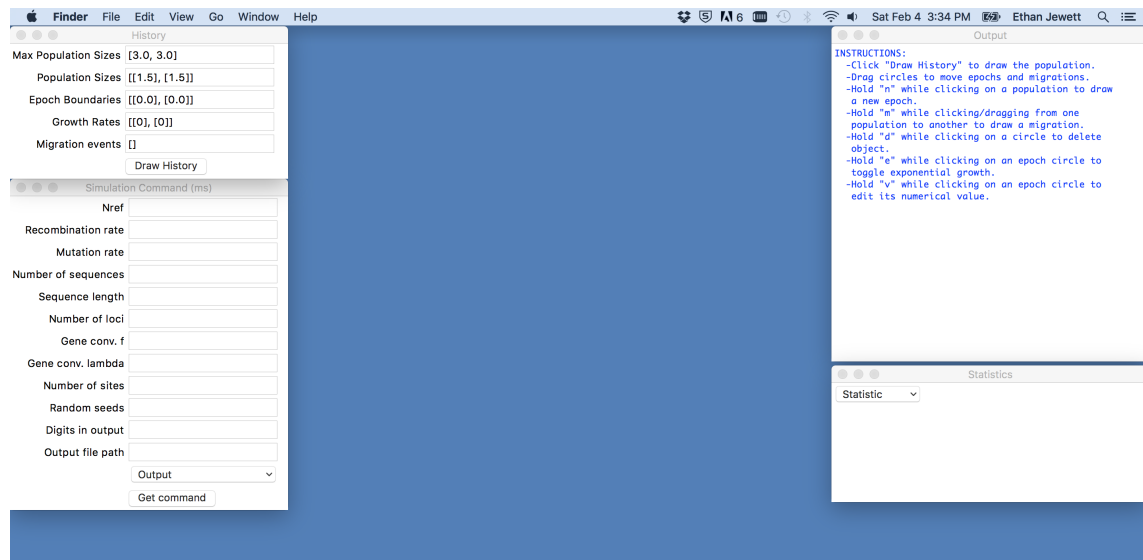


Moving clockwise starting in the top right-hand corner we have

1. **The output text box.** All of the commands you generate will be printed here, along with other user-interaction messages.

2. **The statistics window.** Lets you visualize and compute various population-genetic statistics.

3. **The simulation command window.** Lets you generate commands automatically for different population-genetic simulators.

4. **The history window.** Displays all of the information about your population history as DemoGrapher data objects.

# 3   Drawing the history

To draw the population history and get started, click the "Draw History" button in the
**History** window. You'll get an initialized canvas that looks something like this:
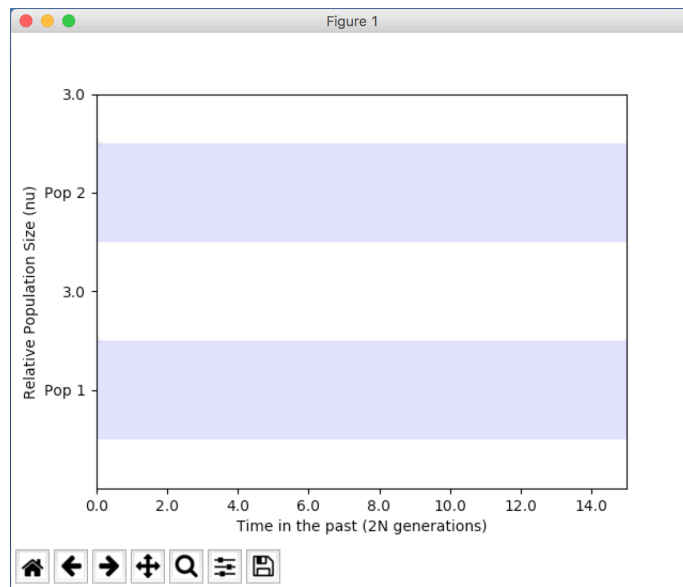


Each blue bar is a population of constant size stretching backward in time from left to
right.

# 4    Drawing population epochs

To draw your population history, hold down the 'n' key on your keyboard and click somewhere in the top population. You'll get something like this:
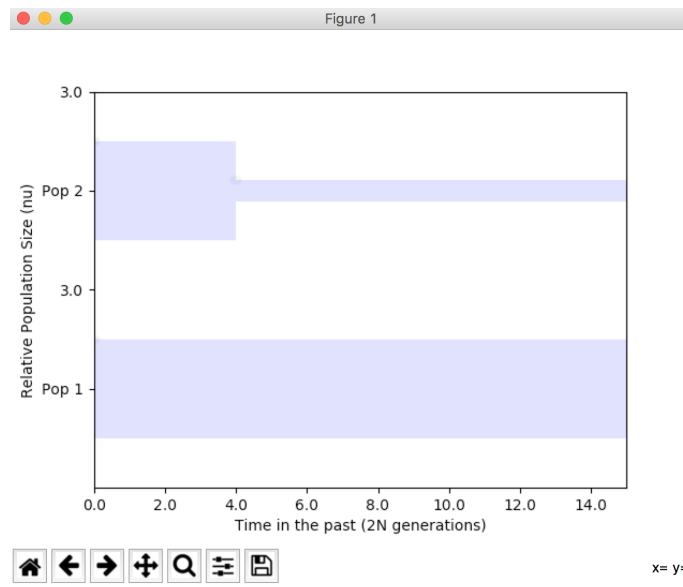


- You just created an second epoch in population 2 that starts around time 4 coalescent units in the past and has a smaller size than that of the first epoch.

- You can drag the little grey circle to move this epoch boundary and size around.

- You can specify the value more precisely by holding down the 'v' key while clicking on the circle. This will open up a dialog that lets you set the size and time values.
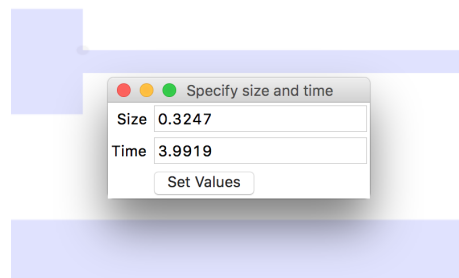
# 5 Exponential growth

To change an epoch to exponential growth, hold down the 'e' key and click the grey circle at the epoch boundary. For example, if you hold 'e' and click the grey circle for the first epoch in population 2, you'll get:
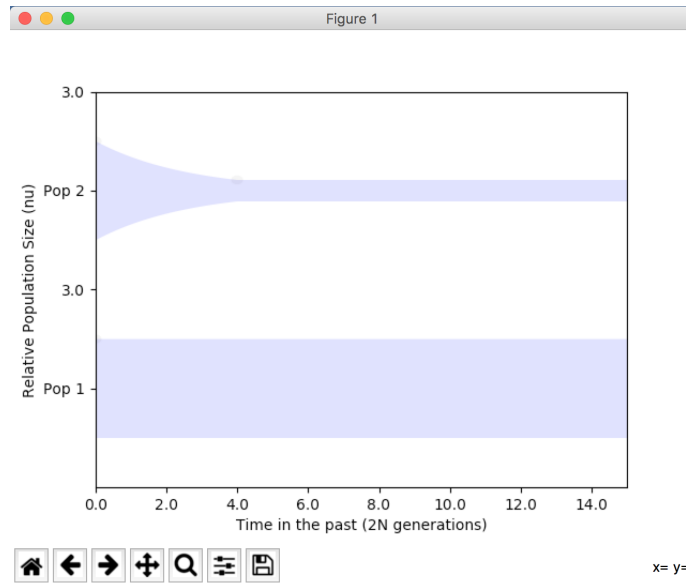
# 6   Migration

To create a migration event from population 2 to population 1, hold down the 'm' key, click down the mouse button in population 2, drag to population 1, and release. You'll get something like:
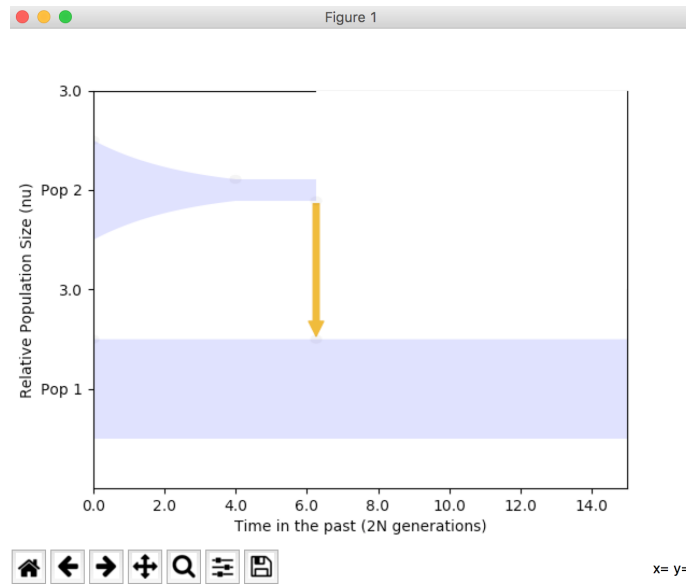


This creates an instantaneous migration event from population 2 to population 1. This is a pulse migration with a strength of 100%, meaning that all the lineages in population 2 were immigrants from population 1 at this time.

- To move the time of the migration event, click and drag its base circle.

- To reduce the migration strength (the fraction of individuals who migrate), drag the tip circle back toward the base.

- To create a continuous migration event, drag the arrow tip circle to the right.

To specify all migration parameters precisely, hold 'v' and click the base or tip circle.

For example, if we hold the 'v' key and click a grey circle on the migration event, we get a popup text box that looks like this.
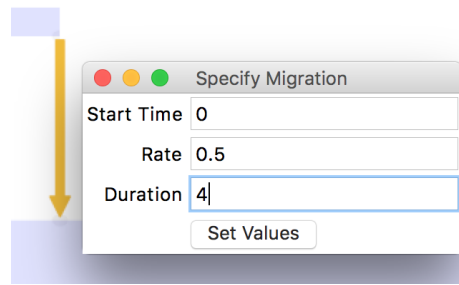


Setting the migration time to 0, the migration strength to 0.5, and the migration duration to 4 gives
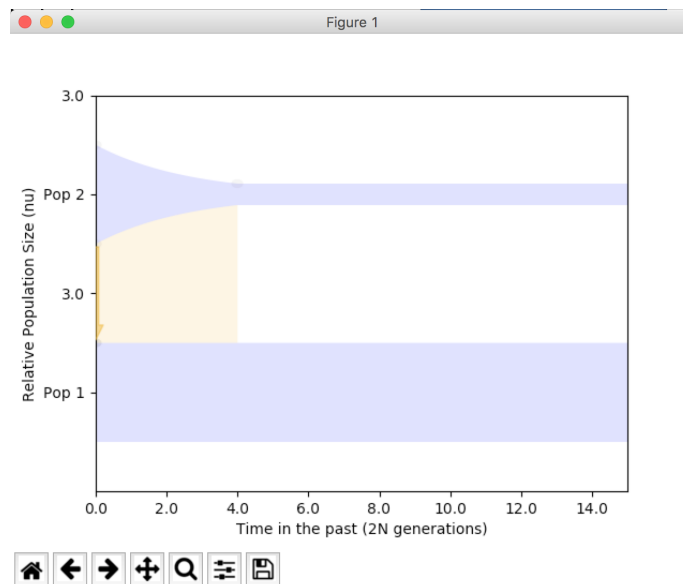
# 7 Computing statistics

You can also use DemoGrapher to compute various statistics. This can be useful as a sanity check before running time-consuming simulations and it can provide some intuition about how population-genetic quantities change under varying demographic histories.

- To compute statistics, mouse over to the **Statistics** window.
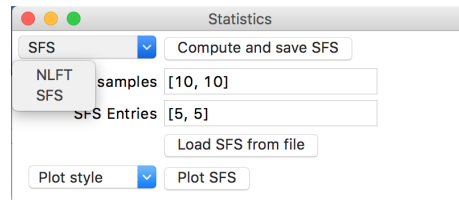


- DemoGrapher currently lets you compute two statistics: the joint sample frequency spectrum (SFS) and the number of ancestral lineages as a function of time (NLFT).

- The SFS can be computed in two ways: 1) using the program *momi* by Kamm et al. (2016) or using a fast asymptotic approximation. *momi* is fast and accurate, but it only works for point migrations. The asymptotic approximation works for both pulse and continuous migration and it can be much faster than *momi* when the sample size is large. However, it can be inaccurate for small samples sizes (like $n \approx 10$).

- To use *momi*, download and install *momi* from `https://github.com/popgenmethods/momi` and then select "Set momi directory" from the **Options** menu. Note: currently you can only compute the SFS with *momi* for populations with pulse migrations.

# 8    Computing the SFS

The SFS a multidimensional array, denoted by $\xi$, with $K$ dimensions, where $K$ is the number of populations. Entry $\xi_{i_1,i_2,...,i_K}$ records the expected number of segregating sites with $i_1$ copies in population 1, $i_2$ copies in population 2, etc. You can either compute the full SFS for all populations, or plot the SFS for one or two populations.

**Computing the SFS**

- The SFS requires you to specify the number of samples in each population and the number of SFS entries to compute.

- For example, for two populations, setting "Number of samples" to [20,15] sets up the computation for an SFS with 20 sampled sequences in Population 1 and 15 sequences in Population 2.

- Setting "Number of samples" to [10,5] computes the upper left $(10+1)\times(5+1)$ corner of the joint SFS matrix consisting of entries $\xi_{ij}$ for $i \in \{0,...,10\}$ and $j \in \{0,...,5\}$.

- To compute the SFS, click "Compute and save SFS."

    - A dialog will open where you can choose the location and file name.
    - The SFS is saved as a *numpy* ndarray in a file with a ".npy" exension.
    - You can load and explore the saved SFS in Python. For example to explore it in iPython, open iPython by typing `ipython` at a terminal prompt. Then import numpy using `import numpy` and load the sfs using `sfs = numpy.load('pathtosfs.npy')`.

**Plotting the SFS**

- Plotting is done in at most two dimensions, so you must choose two populations.

- To do this, set all but two values in the "Number of samples" field to zero.

- For example, if I have four populations and I want to compute the SFS for sample sizes of 25 and 30 in populations 2 and 4, I would type [0, 25, 0, 30] in the "Number of samples" text field.

- Then choose whether you want a heatmap or a line plot from the "Plot style" drop down menu.

- Then press the "Plot SFS" button.

- (WARNING: the asymptotic approximation is currently quite slow for conitinous migration!!! Speed-ups coming soon!).
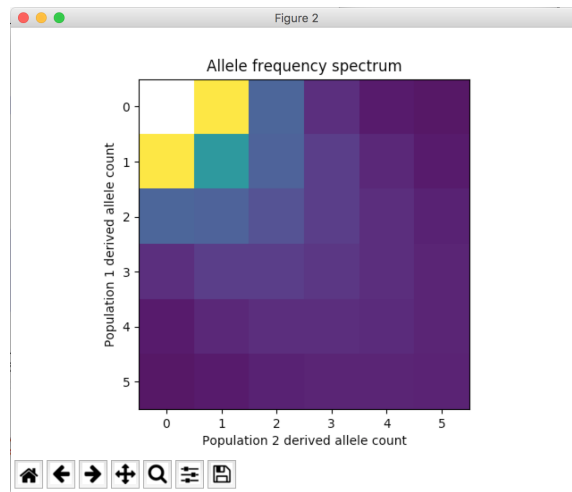
Figure 1: Heatmap of the SFS for the population on page 7 with "Number of samples" set to [50,50] and "SFS Entries" set to [5,5].
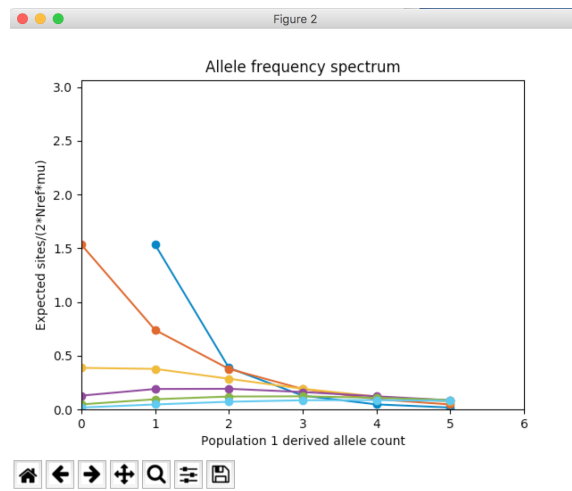


Figure 2: Line plot of the SFS for the population on page 7 with "Number of samples" set to [50,50] and "SFS Entries" set to [5,5].

# 9    Plotting NLFT

The number of ancestral lineages as a function of time (NLFT) can also be visualized with DemoGrapher.

- To visualize the NLFT, choose "NLFT" from the drop-down menu in the **Statistics** window.

- Enter the number of lineages at time zero in each of the two populations in the "Number of samples" text field.

- For example, for a history with two populations, entering [30,25] in the "Number of samples" text field will compute the NLFT when 30 sequences are drawn from population 1 and 25 sequences are drawn from population 2.
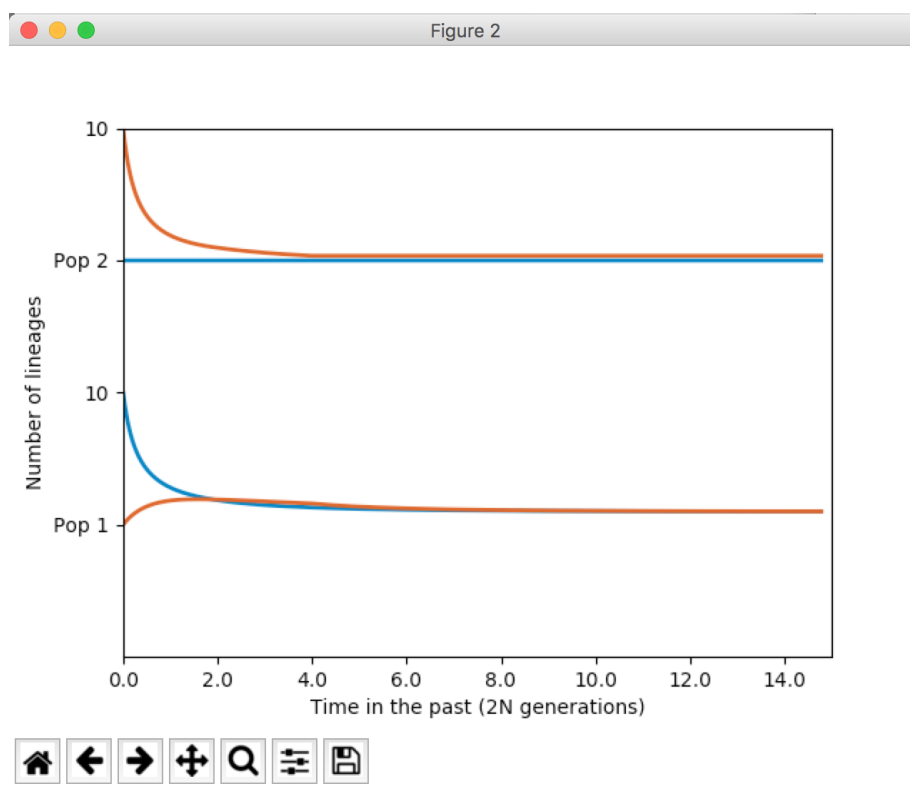


Figure 3: Number of lineages as a function of time (NLFT) for the population on page 7 with "Number of samples" set to [10, 10]. Each curve shows the number of lineages of a given type in each population. For example, the blue curve in population 1 shows the number of ancestors in this population of the 10 lineages that were initially sampled in population 1. The blue curve in population 2 shows the number of ancestors in this population of the 10 lineages initially sampled in population 1 (the migration is all from population 2 to 1, so this curve is zero). Conversely, the red curve in population 2 shows the number of ancestors in this population of the 10 lineages sampled from population 1. The red curve in population 1 shows the number of ancestors in this population of the 10 lineages sampled from population 2.

# 10  Generating simulation commands

You can generate a simulation command directly from your diagram. This provides functionality that is similar to the program PopPlanner by Ewing et al.[4]  Commands can currently be generated for *ms*, *msprime*, and *scrm*. Commands for other simulators are being implemented. To automatically generate a command, mouse over to the "Simulation Command" window (the window for *ms* is open by default).



- To generate a command, enter the required values in the text fields. DemoGrapher provides mouse-over tooltips for all of the parameters so that you don't have to go searching in the *ms* manual. For example, the tooltip for "Number of sites" is shown in the diagram.

- Now click the "Get Command" button. DemoGrapher will automatically generate the command and print it in the **Output** window.



---

[4]Ewing, G., P. Reiff, and J.D. Jensen (2015). PopPlanner: Visually constructing demographic models for simulation. *Front Genet.* **6**: 150.