

BATEMANATER USER MANUAL

by Adam G. Jones

agjones@tamu.edu

Department of Biology, Texas A&M University

College Station, TX 77843

Introduction

This program is intended to solve a couple of problems in the estimation of mating system variables from natural populations. The main problem is that with incomplete sampling of offspring, estimates of variances in mating success and reproductive success will be biased (Mobley and Jones 2013). The Bateman gradient, which is the slope of the regression of reproductive success on mating success, also will be biased. The bias can be removed (or at least reduced) by using a maximum-likelihood approach to estimate these mating system variables. This problem was the motivation for this program. The other problem is that even when samples are complete, confidence intervals are difficult to determine for important mating system variables. A simple solution to this problem is to determine confidence intervals by bootstrapping. Thus, even for samples that include the entire population, this program can be used to determine confidence intervals for the variables of interest. See the key papers listed below for more details. This program only applies to organisms with discrete egg masses that can be collected, either with or without parents, such as frogs, guppies, pipefish, and so on. For organisms in which offspring occur individually and unassociated with parents, the issues addressed by this program have not been solved.

Key Papers

Mobley, K. B. and A. G. Jones. 2013. Overcoming statistical bias to estimate genetic mating system parameters in open populations: A comparison of Bateman's principles between the sexes in a sex-role-reversed pipefish. *Evolution* 67:646-660.

Jones, A. G. 2009. On the opportunity for sexual selection, the Bateman gradient and the maximum intensity of sexual selection. *Evolution* 63:1673-1684.

Using the Program

Part I. Bootstrapping for Confidence in the Sex with Complete Data. Throughout this guide, I assume that the females can be collected with clutches (as in the case of pregnant guppies or some similar organism). Males cannot be collected with their progeny, but their mating success and reproductive success can be at least partially inferred from parentage analysis or direct observation. Thus, all of the mating and reproductive success of each female is present in the clutch collected with her. Under these circumstances, missing clutches do not introduce bias for the female, so it's a simple matter to calculate the mating success and reproductive success variables for females and to estimate confidence intervals. BATEMANATER determines confidence intervals by bootstrapping across females. Here's how to do it in the program:

(1) Launch the program by double-clicking on the icon (if it's zipped, you will need to extract all files first).

(2) Hit the filing cabinet icon on the toolbar or select the menu item "Parameters and Data" from the "Run" menu.

The parameters and data window will pop up. For this application of the program, you can ignore everything except the bottom text box and the number of bootstrap replicates.

(3) Enter the desired number of bootstrap replicates in the box to the right of "Bootstrap Replicates".

For this application, 1000 bootstrap replicates might be a good number, because the program will run very quickly.

(4) Paste your data into the text box (see figure on next page).

Each line should represent a single individual from the data set. Enter the mating success (number of mates), followed by a tab character, the reproductive success (number of offspring) and then a carriage return. The data can be copied and pasted from a text file (but not from a spreadsheet – save it as tab-delimited text and open it with Notepad). Even though the caption above the box indicates that the data for males should be entered in the box, data for either sex can be used, provided the user realizes which type of analysis is necessary for the sex in question. For now, we are dealing with data for females of a species in which pregnant females can be collected, so we merely want to calculate variances in mating success and reproductive success, as well as the Bateman gradient.

(5) Hit "Okay" to shut the data entry window. If you close the window without pressing “Okay”, your data will be discarded.

(6) Click the calculator button on the toolbar or select "Bootstrap Raw Data for CI's" from the "Run" menu.

(7) Look at the results.

The output consists of every bootstrap replicate, plus a summary at the end. For each bootstrap replicate, the program presents means and variances in mating success and reproductive success. It also provides

the opportunity for sexual selection (I_s), which is the variance in mating success divided by the square of the mean, and the opportunity for selection (I), which is the variance in reproductive success divided by the square of mean reproductive success. The Bateman gradient (BateGr) for each bootstrap replicate is reported, along with the standardized Bateman gradient (BG'; regression of relative reproductive success on relative mating success) and s'_{\max} (the square root of I_s times the standardized Bateman gradient; see Jones, 2009, for an explanation of this variable). At the bottom of the output, a table summarizing the results is provided. The first column indicates the variable. The second column gives the point estimate of the variable. The third and fourth columns give the lower and upper 95 percent confidence intervals for each variable. Finally, the fifth and sixth columns give the 90 percent confidence intervals for each variable. If the number of bootstrap replicates is such that it's not possible to calculate 95 or 90 percent confidence intervals, the numbers in parentheses will indicate the actual confidence intervals (as close as possible to 95 or 90 given the number of replicates).

Batemanator Data Entry

Demographic Data

Mean Mating Success of Females: 2.33

Mean Reproductive Success Females: 259.19

Sex Ratio [Males/(Males+Females)]: 0.5049

Estimated Number Males in Population: 436

Number of Clutches Sampled: 63

Mating Success Parameters

Max. Std. Dev. MS (multiple of mean): 3

Number of MS intervals: 200

Bateman Gradient Parameters

Estimation Std. Dev. MS: 2.6

Max. Std. Dev. RS (multiple of mean): 3

Number of RS intervals: 25

Max. Increase RS per mate (x mean): 1.2

Number of Increase Intervals: 20

Estimation Replicates: 250

Bootstrap Replicates: 200

Hypothetical Population Parameters

Increase in RS per mate: 120

St. Dev. in RS within mating class: 250

Mating System Data

Enter Results of Parentage Analysis on Males Here. Each row should correspond to a male. First Enter his mating success, followed by a tab and his reproductive success (i.e., 4[tab]231[Enter]).

1	153
1	173
1	219
1	223
1	247
1	253
1	408
5	471
2	147
2	118
2	294
3	381
3	726
4	499

Okay

Set number of bootstrap replicates here

Enter data here

Part II. Obtaining an Estimate of the Mating Success Distribution for a Sex with Missing Clutches. If we continue with the example in which pregnant females can be collected with their clutches, then the real challenge lies in determining the mating success distribution of the males. If the entire population of clutches can be collected and parentage can be assigned with complete certainty, then there is no problem and the variables for both sexes can be estimated using the procedure in Part I. If, however, some clutches are missing from the dataset, then we need to take care in estimating the mating system variables of interest (as discussed in Mobley and Jones 2013). To run the program successfully, you will need estimates of the following variables:

(a) The mean mating success of females: Usually this value is just the average mating success of all the females for whom you have data. Remember that for these females you have complete data because they were collected with their entire clutches.

(b) The mean reproductive success of females: This value should just be the average number of offspring across all the females included in your dataset. This average should be their average total number of offspring, not their average number sampled or assayed.

(c) The sex ratio: This value should be expressed as the proportion of adults in the population that are male.

(d) The estimated number of males in the population: This value should be an estimate of the total number of adult males in the local breeding population (i.e., the number of potential sires that could have mated with the females in your dataset). If you have an estimate of the local breeding population size, then this value should be the population size times the sex ratio. There are ways to estimate the number of males using a genetic “mark-recapture” approach.

You can ignore all of the other parameters in the Batemanater Data Entry window if you wish!

The following steps will allow you to run the analysis:

(1) Enter the Mean Mating Success of Females, Mean Reproductive Success of Females, Sex Ratio, and Estimated Number of Males in Population under Demographic Data in the Batemanater Data Entry Window.

(2) Enter your mating system data in the large window near the bottom of the Batemanater Data Entry window.

See number (4) in Part I for the format of the data. These data should be the values for the **males**. Clearly, these data will be incorrect because they will suffer from the bias described by Mobley and Jones (2013). Nevertheless, these data serve as the basis for the estimation procedure. In other words, these are the apparent mating success and reproductive success of males that appeared in your data set (i.e., inferred from the broods of the females). **There should be no zeros in these data**, because you cannot infer zero mating success for any males (they may have mated with females you did not collect). Even the means will be wrong, probably much lower than the means for the females, but the program will correct these, too.

(3) If you intend to bootstrap, set the number of bootstrap replicates as shown in the figure above. These procedures will take a long time to run, so you might want to choose a modest number like 200 initially. For your final analysis, you will probably want to use something like 1000 replicates.

(4) Press Okay.

(5) Push the alarm clock or “Estimate Mating Success Distribution” menu item.

This procedure will run through all of the allowable distributions of mating success, calculating an exact probability of the data for each distribution (see Mobley and Jones 2013 for the formula). The distribution with the highest probability of producing the data will be considered the maximum likelihood estimate. The program reports the mean male mating success, the standard deviation in mating success and the opportunity for sexual selection. It also reports the “Std. Dev. for Simulation”, which has no meaning to anyone but the program. This value is used in the estimation of the standard deviation in reproductive success and the Bateman gradient, and the program will automatically enter it for you into the appropriate place in the Data Entry window (“Estimation Std. Dev. MS”). If for some reason, you ran “Estimate Mating Success Distribution” and then closed the program, you would have to enter this value manually before you tried to estimate the standard deviation in reproductive success and the Bateman gradient. Only this parameter is passed on from this step to the next one, however.

There are two parameters related to “Estimate Mating Success Distribution”:

(a) Max. Std. Dev. MS (multiple of mean): This parameter sets the maximum allowable standard deviation in mating success. It is determined as a multiple of the mean, so the maximum value for the standard deviation is this parameter value times the mean male mating success. In general, the default value of three will be high enough for most mating systems.

(b) Number of MS intervals: This parameter determines the number of steps the program takes to get from the lowest value to the highest value for the allowable values of the standard deviation in mating success. Higher values will result in more precise estimates at the cost of a longer run time.

(6) Push the Earth button or select “Estimate St. Dev. RS and Bateman Grad.” from the Run menu.

If the estimation of the mating success distribution seemed to work, now you can estimate the standard deviation in reproductive success and the Bateman gradient. You must have an estimate of the “Estimation Std. Dev. MS”, which is automatically estimated by the previous procedure. It is also automatically entered into the appropriate place for you.

After this procedure runs, the program will report, for males, estimates of the Mean Mating Success, the Mean Reproductive Success, the Standard Deviation in Mating Success, the Standard Deviation in Reproductive Success, the Bateman Gradient, the Bateman Gradient without the zero class included (lots of zeros will tend to force it through the origin, but I still recommend keeping the zeros in the analysis), the standardized Bateman gradient (that is, relative reproductive success on relative mating success), the opportunity for sexual selection, the opportunity for selection, and s'_{\max} . Two other values are reported, and these are used by the program and have no real meaning to the user.

(7) Perform the whole procedure with bootstrapping by pushing the hand pointing to the left or “Estimate and Bootstrap St. Dev. MS, RS and BG” from the Run menu.

If the program gave you seemingly reasonable answers when you ran steps 5 and 6, then you may choose to embark on the bootstrapping procedure, which will take a while. Set the number of bootstrap replicates to the desired level, make sure your parameters are entered, and press the button. Ideally, you want to use the “Estimate and Bootstrap St. Dev. MS, RS and BG” routine, because it performs the entire estimation procedure starting with the estimation of the mating success parameters all the way through the estimation of reproductive success variance and the Bateman gradient. Thus, any uncertainty in the estimation of the mating success distribution will contribute to the confidence intervals of the estimates of the variance in reproductive success and the Bateman gradient, as it should.

To run this step, you need only enter the Demographic Parameters and your data, as described in Steps (1) and (2) above in Part II.

Some other considerations:

(A) Choose the “Estimated Number of Males in Population” parameter carefully. Because the estimation of reproductive success and the Bateman gradient involves simulating the entire population, large values here will substantially slow down the simulation. As a rule of thumb, I would not set this parameter larger than about 2000 even if the population is extremely large.

(B) What if you have data on mating success but not reproductive success? Obviously, you cannot estimate the variance in reproductive success, but you can still estimate the distribution of mating success. You will need to put arbitrary dummy values in for each individual’s reproductive success (the data format will still be the same, but the reproductive success values are not used at all when calculating mating success variables). Then just run either “Estimate Mating Success Distribution” for a point estimate or “Estimate and Bootstrap MS Dist.” for an estimate with bootstrapped confidence intervals.

(C) Use “Save Log File” if you think your computer might crash. If you check this checkbox, then the program will ask you for a folder location. The program will then save the bootstrap results every five replicates, and these results can be used to manually calculate confidence intervals if necessary. At the end of the run, the program will also save a results file with the results. Many files will be generated (one for every five bootstrap replicates), so select a folder that you don’t mind filling up with many very small text files.

(D) Push the ambulance button to output a hypothetical distribution of mating and reproductive success based on the estimated distributions of mating success, reproductive success and the Bateman gradient. This button simulates a dataset that would be obtained from a hypothetical population with the estimated mating system values if knowledge of the mating system had been complete. In other words, this button shows you what the program thinks the “real” males in the population would have had in terms of total mating and reproductive success. These data can be useful to visualize what the population’s actual Bateman gradient might have looked like (given all the assumptions that go into the estimation procedure). This procedure requires several parameters: “Estimation Std. Dev. MS”, “Increase in RS per mate”, and “St. Dev. in RS within mating class”. These

parameters will be entered in their appropriate boxes automatically by the various subroutines. However, if you just turned on BATEMANATER, then you will need to find these values in your output files and enter them yourselves in the “Batemanater Data Entry” window.

(E) The light bulb button just clears the text screen.

(F) Save your results using “Save”. This button just saves the contents of the text window, including anything you decide to type in there.

(G) Estimate and Bootstrap MS Dist. This procedure estimates only the mating success distribution, so use it if you’re not interested in variance in reproductive success or the Bateman gradient.

(H) Estimate and Bootstrap St Dev RS and BG. This button estimates the standard deviation in reproductive success and the Bateman gradient with bootstrapping. You must have an estimate of the “Estimation Std. Dev. MS” parameter entered in the appropriate window in the Batemanater Data Entry window for this procedure to provide meaningful results, however. It’s also preferable to do the entire analysis in one massive run, as described above, so in general you are not going to want to use this button.

(I) What are all those other parameters under “Bateman Gradient Parameters”?

(a) Max. Std. Dev. RS (multiple of mean): This parameter sets the maximum standard deviation in reproductive success that the program will test. The default is a standard deviation three times larger than the mean.

(b) Number of RS intervals: This parameter sets the number of steps that the program uses to get from the lowest value of the standard deviation in reproductive success (zero) to the maximum value. More RS intervals will increase precision but slow run times.

(c) Max. Increase RS per mate (x mean): This parameter sets the maximum slope of the Bateman gradient. Because the standardized Bateman gradient should never be larger than one, in principle, a value of 1.2 here should be high enough for most purposes.

(d) Number of Increase Intervals: This parameter sets the number of intervals the program uses to step from the lowest slope to the highest slope for the Bateman gradient. Higher values result in more precise estimates but slow the analysis.

(e) Estimation Replicates: The procedure used by BATEMANATER to estimate the standard deviation in reproductive success and the Bateman gradient simulates breeding populations and samples from them to compare the simulated population to the actual observed sample. More simulations can improve the accuracy of estimates, in principle, but they take time. The default is to simulate 250 population for each parameter combination. Higher values could result in better results in some cases.

(J) What about bugs or questions? If you experience any bugs or have questions, email Adam Jones at agjones@tamu.edu.