M & M evolution by drift

Simulating genetic drift of a haploid organism using M & Ms

AUTHOR
BIO 107 lab

Instructions

1. You will create a google sheet for this lab assignment. Please share the google sheet with your lab instructor by 1) click the share button in google sheets and add your instructor's e-mail address, and 2) **IMPORTANT** paste the URL to the google sheet in the assignment submission page in Brightspace!
2. Answer the questions on this sheet and turn in to instructor at the end of the lab

Background knowledge

Questions to answer

Q1. what is an allele?

Q2. what does haploid mean?

Q3. what is an example of a haploid organism?

Q4. are human cells haploid? Explain.

Q5. what is the difference between an absolute and a relative frequency of alleles? Give me an equation for the relative frequency

Simulation 1: Genetic drift with initial population size of $n = 20$

(in groups of four). In this exercise, you will simulate an evolving, haploid population of M & Ms. The color of an M & M represents a different allele at a locus (a gene) on the M & M genome. Each generation, the pool of alleles in the population will be sampled and duplicated. You will track the **absolute frequency** of the alleles in the population through time.

Preparation

To start, each group of 4 students should divide into two groups of two. Each subgroup will do the following.

3. Create a new Google Sheet, titled "BIO 107 lab m & m - Last names"
4. title the sheet "drift (n = 20)"
5. In cells A1:G1, write the column labels "generation", "red", "blue", "yellow", "orange", "green", "brown", "total".
6. insert the number *0* in the cell A2. Then, insert the function "=A2 + 1" in cell A3. Then, fill down to make a column with the numbers 0:10.
7. insert the formula to sum the number of each color M & M in cell H2 and fill this down.

You will record the **absolute** frequency of each color M & M each generation. The colors are the alleles in the population. Refer to the colors as the "red allele" or "blue allele"

Now...

8. Create a 2nd sheet and title it "random sampler"
9. Create three column headers in cells A1, B1, and C1:
    1. "id" – this is the id of your m & m in the population (the position in the row)
    2. "prob" – this is the probability of the m&m reproducing
    3. "reproduce" – this is the random event, 1 = reproduce, 0 = does not reproduce

10. insert the number *1* in the cell A2. Then, insert the function "=A2 + 1" in cell A3. Then, fill down to make a column with the numbers 1:20
11. insert the the number *0.5* in cell B2. Then fill down so that each cell has *0.5*
12. insert the function "=binom.inv(1, B2, rand()) in cell C2. Then, fill

down so that each cell has this function.

The binom.inv function is sampling from the **binomial distribution** with a single sample and probability of the number in cell B2, which should be 0.5. This is a coin flip. The result is a number of 1, which you can think of as "success" (reproduces!) or a 0, which you can think of as "no success" (no reproduction).

Finally,

13. On a blank piece of paper, write the numbers 1:50 spaced far enough to position an M & M above each. These numbers will be the M & M ID used in the "random sampler" sheet of the Google sheet.

	A	B	C
1	id	prob	reproduce
2	1	0.5	0
3	2	0.5	1
4	3	0.5	0
5	4	0.5	1
6	5	0.5	1

Simulate genetic drift with M & Ms

14. blindly pull 20 M & M pieces from the bag. This is generation 0. Record the absolute frequencies of each allele for generation 0 in the "drift (n = 20)" sheet.
15. line up the M & Ms on your numbers on your sheet of paper. The number is the ID of the M & M.
16. refresh your random sampler and use the results to identify which M & M reproduces and which do not.
17. separate the M & Ms that reproduce and return the others to the bag
18. duplicate each of the M & Ms that reproduce by matching each parental M & M with the same-color daughter M & M from the bag.
19. record the new frequency of each allele in the row representing the next generation on the spreadsheet in the sheet "drift (n = 20)"
20. repeat these steps, starting with "line up the M & Ms on your numbers", until generation 10 (so you should have 11 generations of data in total)

Notes

21. if your population of M & Ms grows above 30, then add IDs to your spreadsheet and numbers to your piece of paper to accommodate this!
22. if your population of M & Ms goes extinct before generation 10, you're done! This really happens!

Plot the evolution of M & M alleles

Use the charting feature of Google Sheets.

23. Highlight the matrix of data including
    - the column including generation number
    - the row including the column labels

but do not include in the highlight

```
* the column "total"
```

24. Choose $Insert > Chart...$ from the menu
25. Click **Chart Types**
26. Choose the 1st (left most) line chart (the icon with an angular red and angular blue line)
27. Click the Insert button.

Questions to answer

Q6. Is the starting frequency (generation 0) of the different alleles the same in subgroup 1 and subgroup 2? Why or why not? What evolutionary process does this simulate (check your textbook).

Q7. Is the pattern of change in allele frequency of the alleles in the subgroup 2 experiment the same as the patterns for the same alleles in the subgroup 1 experiment. Why or why not?

Q8. Do any of the alleles go extinct?

Q9. In general does genetic drift increase genetic diversity, decrease genetic diversity, or have no effect on genetic diversity?

Q10. Which alleles increased in frequency from generation 0 to generation

10?

Q11. If any alleles increased in frequency, which of these, if any, did you have a prior expectation that it would increase in frequency. Why or why not?

Simulation 2: Initial population size of $n = 40$

Replace all M & Ms in the bag and repeat the simulation with an initial population size $n = 40$.

28. Create a new sheet titled "drift (n = 40)"
29. Add extra rows in the n = 40 sheet as needed

Plot the evolution of M & M alleles for the $n = 40$ simulation

As above

Questions to answer

Q12. did any of the 12 alleles (6 in each replicate subgroup) go extinct? If so, how many?

Q13. is this more or fewer than went extinct in the $n = 20$ simulation?

Q14. would you expect more alleles, or fewer alleles, or about the same number of alleles to go extinct in the $n = 40$ experiment? Why?

Q15. In general does genetic drift increase genetic diversity, decrease genetic diversity, or have no effect on genetic diversity? How does population size effect this?

Simulation 3: Selection!

30. Create a new sheet titled "selection (n = 20)", this should look like the "drift (n = 20)" sheet
31. On the "random sampler" sheet, add the column labels in cells E1:H1 (this leaves a blank column in between this net set and the old set)
    - "prob red" - this is the probability of a red M & M reproducing
    - "prob not red" - this is the probability of any non-red M & M reproducing
    - "reproduce red" - this is the random event **if the M & M with the ID is red**, 1 = reproduce, 0 = does not reproduce
    - "reproduce not red" - this is the random event **if the M & M with the ID is not red**, 1 = reproduce, 0 = does not reproduce

32. Use 0.6 as the probability for a red M & M
33. Use 0.4 as the probability for a non-red M & M
34. Copy and paste the formulas for the binomial probability. **Make sure to modify the cell that is pointed to to insure the correct probability for that color M & M!**

Now, replace all M & Ms in the bag and repeat the n = 20 simulation but using the random sampler with different sampling probabilities for red and non-red M & Ms

Plot the evolution of M & M alleles for the selection simulation

As above

Questions to answer

Q16. Which alleles increased in frequency from generation 0 to generation 10?

Q17. If any alleles increased in frequency, which of these, if any, did you have a prior expectation that it would increase in frequency. Why or why not?

Q18. did any of the 12 alleles (6 in each replicate subgroup) go extinct? If so, how many?

Q19. is this more or fewer than went extinct in the $n = 20$ drift simulation?

Q20. would you expect more alleles, or fewer alleles, or about the same number of alleles to go extinct in the *this* selection simulation compared to the $n = 20$ drift simulation? Why?