ACTIVITIES for students

Sarah I. Duncan, Suzanne Lenhart, and Kelly K. Sturner

Measuring Biodiversity with Probability

JAMES DEBOER/THINKSTOCK

More and more, teachers are asked to make connections between the STEM disciplines—science, technology, engineering, and mathematics—while also addressing state standards. Mathematics is an underappreciated but important tool for the life sciences, from mathematically modeling biological processes to making sense of biological data. The activity presented here was designed for a Girls in Science camp, held at Tremont, Tennessee, in the Great Smoky Mountains National Park. The camp is designed to give local girls entering eighth grade a chance to become familiar with the natural world by doing hands-on research in the park. This particular exercise was designed to show the value of mathematics for quantifying and interpreting biodiversity data that the girls had collected.

Biodiversity is important for ecosystem health and productivity. Greater biodiversity provides greater opportunities to find new organisms that can be important food sources or new medicines for people around the world. Also, high biodiversity makes ecosystems more resilient; with increasing threats to species from global climate change or natural catastrophes, more species can mean a greater potential that some organisms will have the necessary traits to survive.

But what exactly is biodiversity? How do we go about measuring whether biodiversity in one place is greater than biodiversity in another? Scientists and ecologists do not simply count up all the different organisms and compare the totals. Measuring biodiversity involves applying probability to obtain a more meaningful index for comparison.

These two activities allow students to apply tools from mathematics to life science, raising awareness of connections between STEM disciplines. The activities were developed for grades 9–12 and will take approximately one hour each, depending on the students' level of experience with probability. Mathematics teachers using these activities are encouraged to communicate or work with biology teachers to strengthen the activity. If time allows, students could collect data and talk about biodiversity as part of their biology class exercise and complete the activity in a mathematics unit on counting or probability. Group work is also appropriate for this exercise to allow students to discuss difficult terminology and reduce the amount of class time the activity will take.

Activities for Students appears six times each year in *Mathematics Teacher*, often providing in reproducible formats activity sheets that teachers can adapt for use in their own classroom. Manuscripts for the department should be submitted via http://mt.msubmit.net. For more information on the department and guidelines for submitting a manuscript, please visit http://www.nctm.org/publications/content.aspx?id=10440#activities.

Edited by ***Ruth Dover***
dover@imsa.edu
Illinois Mathematics and Science Academy
Aurora, IL

Patrick Harless
pdharless@gmail.com
University of Rochester (PhD student)
Rochester, NY

Activity 1 serves as an introduction to quantifying biodiversity and can be used alone for students in beginning high school mathematics courses. This activity will be most successful in a classroom in which students have a basic understanding of counting and probability. It also provides a real-world application for these skills. Students are introduced to Simpson's Diversity Index, which uses principles of probability to determine objectively which area has the greater biodiversity. After being led through a derivation of Simpson's Diversity Index, students then apply the equation to a data set on their own. At the end, students may discuss whether their results support their original hypotheses. Activity 1 provides an opportunity to learn and use knowledge connected with the following Common Core State Standards: seeing structure in expressions (algebra), building functions (functions, modeling), and conditional probability and rules of probability (statistics and probability, modeling) (CCSSI 2010).

For advanced classes (usually precalculus or higher), the module may be continued with activity 2, which requires an understanding of how to calculate limits. Students use mathematical practice skills from the Common Core State Standards—reasoning abstractly and quantitatively, constructing viable arguments, and using appropriate tools strategically (CCSSI 2010)—to discover how species richness and species evenness affect the index. Considering extreme cases of richness and evenness that may not occur in a biological scenario can aid students in understanding the constraints of this model.

This module may serve as an introduction to measuring biodiversity and then may be applied to a real-data set gathered by students. Students could, for example, count insects from two different areas on the school grounds or learn to inventory tree species found in their yards at home or in a local park.

First, students will learn key vocabulary and concepts. Teachers might begin by asking students for their definition of biodiversity. Already students may implicitly understand that what they are describing is a measurement and involves quantification. The sample data set found in **table 1**, tree species that were counted in two yards, may be used to introduce and illustrate the key concepts of species richness and species evenness.

Table 1 Number of Trees

Species	Yard A	Yard B
Eastern redbud	3	5
Black oak	4	5
Post oak	5	5
White pine	3	5
Honey locust	1	5

- *Biodiversity* is a measure of the different kinds of organisms (species) in a region or other defined area. Usually when quantifying biodiversity, we look at a group of similar organisms (such as all insects), rather than all the organisms in the ecosystem. This term takes into account both species richness and species evenness.
- *Species richness* is the number of unique species in a region or specified area. The data set shows five different species of trees found in the back yard (see **table 1**). Five is a measurement of the richness value.
- *Species evenness* is the degree of equitability in the distribution of individuals among species. Greater evenness signifies less variation in the numbers of individuals of each species. Maximum evenness occurs when the number of individuals among all species is the same. In **table 1**, we see that yard B has exactly the same number of each kind of tree and, thus, has greater species evenness than yard A.

Although richness makes intuitive sense for measuring biodiversity, students may wonder why species evenness would be important. Consider the example of a forest where introduction of a new species gives rise to a beetle infestation that causes population decline in other insect species. One additional species, adding to the richness value of insects but reducing the evenness, would thus have a negative impact on the biodiversity.

A useful next step might be to assess students' understanding of the concepts informally. Teachers can do so by presenting another table of data and having students discuss species richness and evenness.

After students are comfortable with these concepts, ask them to imagine this scenario: An ecologist collects information from two separate experimental plots of the same size but with one big difference: Plot 1 is in the woods, and plot 2 is in a nearby field. The ecologist is interested in the types of insects that are found in the plots. The first data table in activity 1 is an example of the data that might be collected. Before being shown the table, students should think for a minute about what they expect to find and formulate a hypothesis.

A UNIVERSAL WAY TO DETERMINE BIODIVERSITY

Students might wonder which would be considered more biologically diverse—a plot with more richness or a plot with more evenness. What happens if one plot has many more species—would it be as easy to determine which plots had more evenness?

To answer such questions, in 1949 the British statistician Edward H. Simpson devised an objective method to measure biodiversity that is still used today. Simpson suggested that ecologists calculate overall diversity in terms of probability.

Simpson's Diversity Index can be explained as follows: If we randomly select an individual from a sample without replacement and then randomly select another individual from the same sample, what is the probability that the two organisms will be of different species? The higher the probability of different species, the higher the diversity. Simpson's Diversity Index, D, can be represented as

$$D = 1-\left(\frac{n_1\left(n_1-1\right)+n_2\left(n_2-1\right)+\cdots+n_S\left(n_S-1\right)}{N\left(N-1\right)}\right).$$

where S represents the number of species, n_i represents the number of individuals in the ith species, and N represents the total number of individuals. [The quantity 1-D is also sometimes referred to as Simpson's Index, so careful reference to other resources is warranted. D as given here is also known as the Gini-Simpson Index.—*Ed*.]

MORE TO EXPLORE

Biodiversity can be quantified in more than one way. Consider having students do research on Shannon's Index (Shannon 1948), an alternative method of calculating a type of biodiversity index, and compare it with Simpson's Diversity Index. Also, Simpson's Diversity Index is not as useful when representing biodiversity across trophic levels (such as predators and prey). It does not make ecological sense to expect the same population sizes among predators and prey, if one predator needs to consume multiple prey species to survive. Ask students to think about ways to quantify total ecosystem biodiversity. There is no single right answer, but perhaps one student will publish a possible solution in a scientific journal someday.

For a small real-data set to explore, consider **table 2**, showing some of the species of salamanders found in the Great Smoky Mountains National Park. Thirty-one species can be found within the park's boundaries (National Parks Conservation Association, n.d.); thus, the Smokies are often called the salamander capital of the world.

Table 2 Number of Salamanders at Selected Locations in the Great Smoky Mountains National Park

Species	Lower Dorsey Stream	Pig Pen Stream
Spotted dusky salamander	7	18
Imitator salamander	6	3
Seal salamander	5	15
Black-bellied salamander	7	11
Desmognathus spp. salamander	4	17
Blue Ridge two-lined salamander	1	31
Spring salamander	2	1
Northern slimy salamander	0	1
Black-chinned red salamander	0	0
Santeetlah salamander	1	0
Southern red-backed salamander	2	0

Note: These salamanders were identified in submerged bags filled with leaf litter from two different streams (Lower Dorsey and Pig Pen) in 2008. Data were collected as part of a citizen science program at the Great Smoky Mountains Institute at Tremont. After your students explore these data, have them look for more data on the Web or ask them to collect their own.

MATHEMATICS AND BIOLOGY

Integration of mathematics with biology can illustrate that the concepts learned in mathematics class are not isolated exercises but are used in real-world applications that contribute to scientific knowledge. Students may realize that quantifying biodiversity could affect conservation efforts and ecosystem management. Biology-based exercises can provide an exciting and relevant way to reinforce mathematical concepts and explore the natural connections between mathematics and science.

ACKNOWLEDGMENTS

The authors thank the Great Smoky Mountains Institute in Tremont, Tennessee, for allowing them to pilot quantifying biodiversity exercises during their camps for students and teachers and its citizen science program for permission to use a subset of its data. Also, thanks to Jerry Dwyer and Cathy Kessel for their helpful comments on an earlier draft of this article. This work was conducted at the National Institute for Mathematical and Biological Synthesis, an Institute sponsored by the National Science Foundation, the U.S. Departmentof Homeland Security, and the U.S. Department of Agriculture through NSF grant no. EF-0832858, with additional support from the University of Tennessee–Knoxville.

REFERENCES

Common Core State Standards Initiative (CCSSI). 2010. Common Core State Standards for Mathematics. Washington, DC: National Governors Association Center for Best Practices and the Council of Chief State School Officers.

National Parks Conservation Association. n.d. Great Smoky Mountains National Park. http://www.npca.org/parks/great-smoky-mountains.html

Shannon, C. E. 1948. "A Mathematical Theory of Communication." *The Bell System Technical Journal* 27: 379–423 and 623–56.

Simpson, Edward H. 1949. "Measurement of Diversity." *Nature* 163: 688.



SARAH I. DUNCAN, duncansarah3@gmail.com, is a PhD candidate in the Department of Biological Sciences, Ecology, Evolution, and Systematics at the University of Alabama at Tuscaloosa. **SUZANNE LENHART,** lenhart@math.utk.edu, is a Chancellor's Professor in the Department of Mathematics and is the associate director for education and outreach at the National Institute for Mathematical and Biological Synthesis (NIMBioS) at the University of Tennessee-Knoxville. **KELLY K. STURNER,** ksturner@nimbios.org, is education and outreach coordinator at NIMBioS. She is also a graduate student in science education at the University of Tennessee-Knoxville.

more4U

These activity sheets are available to teachers as Word documents that may be copied and edited for classroom use; go to **www.nctm.org/mt**. For solutions to the activities, download one of the free apps for your smartphone and then scan this tag to access **www.nctm.org/mt052**.

ACTIVITY 1: Behind Simpson's Diversity Index

Look at the data below and make a hypothesis. Do you think that biodiversity will be higher in the woods or in the field? Why? Can you suggest a way to decide by using mathematics?

Species	Plot 1: Woods	Plot 2: Field
Pillbug	50	10
Monarch butterfly	36	50
Seven-spotted lady beetle	35	0
Western honeybee	55	39

Simpson's Diversity Index was invented by a British statistician in 1949 and is based on probability. To see how it works, you can derive the equation for the index. We will start with a very simple case:

Imagine that you went into your back and front yards and counted the insects in each area. You found only two species of insects—the Western honeybee and the seven-spotted lady beetle (see chart below).

Insects	Back Yard	Front Yard	Totals
Western honeybee	10	35	
Seven-spotted lady beetle	50	40	
Totals			

You have totals of 60 individuals in the back yard and 75 individuals in the front yard.

1. *Using the back yard data:* Imagine that you put all those insects from the back yard into a bag. Suppose that you reach into the bag and randomly remove an individual and then randomly remove a second individual without replacement.

   (a) What is the probability that the two selected individuals are both Western honeybees?

   (b) What is the probability that the two selected individuals are both seven-spotted lady beetles?

   (c) What is the probability that the two individuals are from different species? Hint: This event is the complement of the union of the two events in questions 1(a) and 1(b).

2. *Using the front yard data:* Imagine that you put all those insects from the front yard into a bag. Suppose that you randomly select an individual and then randomly select another individual without replacement.

   (a) What is the probability that the two selected individuals are both Western honeybees?

   (b) What is the probability that the two selected individuals are both seven-spotted lady beetles?

   (c) What is the probability that the two selected individuals are from different species?

3. Suppose that we now have n_1 of one species and n_2 of a second species in our area. The total number of individuals is now $N = n_1 + n_2$. See the chart below:

Species Type	Number of Individuals
Species 1	n_1
Species 2	n_2
Total Individuals	$\boldsymbol{N}$

Suppose that you randomly select an individual and then randomly select another without replacement.

(a) What is the probability that the two selected individuals are both species 1?

(b) What is the probability that the two selected individuals are both species 2?

(c) What is the probability that the two selected individuals are from different species?

4. Suppose that you have a very large data set, with many more than just 2 species. You might also have monarch butterflies, craneflies, praying mantises, and other insects, for a total of S species (see the chart below).

Species Type	Number of Individuals
Species 1	n_1
Species 2	n_2
Species 3	n_3
$\vdots$	$\vdots$
Species S	n_S
Total Individuals	$\boldsymbol{N}$

Show that the probability that the two selected individuals are different species would be

$$D = 1 - \left(\frac{n_1(n_1 - 1) + n_2(n_2 - 1) + \cdots + n_S(n_S - 1)}{N(N-1)} \right).$$

This is Simpson's Diversity Index, D. Hint: Use your reasoning from question 3(c).

5. If D is close to 1, is the area's biodiversity high or low? Justify your answer.

6. Calculate D for the plot 1 (woods) and plot 2 (field) data. Which plot has the higher D value? Does this calculation support your original hypothesis?

7. Could biodiversity be quantified in other ways? Justify your answer. Make suggestions about other ways to quantify biodiversity. Check the Web to find other indices that give alternative methods of calculating a biodiversity index.

ACTIVITY 2: Exploring the Effects of Species Richness on Simpson's Diversity Index

$$D = 1 - \left(\frac{n_1(n_1 - 1) + n_2(n_2 - 1) + \cdots + n_S(n_S - 1)}{N(N-1)} \right)$$

$$= 1 - \frac{\sum_{i=1}^{S} n_i(n_i - 1)}{N(N-1)}$$

Simpson's Diversity Index (D) takes into account both species evenness and species richness. *Species richness* measures the number of unique species in a region or specified area. *Species evenness* is the degree of equitability in the distribution of individuals among species. But just how important is each factor in determining D? Can you get a very high index value with just one of these factors? In this activity, we will explore these questions to better understand how this mathematical model of biodiversity works. Note that the cases below with the same number of individuals in each species would not occur in nature. We are considering only the case of complete evenness.

1. Suppose a plot has four species with 100 individuals of each species (see the chart at right). Calculate D for this example.

2. Repeat exercise 1 to calculate D with 200 individuals in each species. Then calculate D with 500 individuals in each species. How do these answers compare with each other and with the answer from exercise 1? For $S = 4$, how does D change as we increase the total number of individuals but keep the proportions the same?

Species Type	Number of Individuals
Species 1	100
Species 2	100
Species 3	100
Species 4	100
Total Individuals	**400**

3. Suppose that a plot has four species and n individuals in each species. Calculate D for this example. Show that as we increase n but keep the population proportions the same, the values of D will decrease.

4. Suppose that a plot has S species with n individuals in each species. In this case, find D as a function of S and n and simplify as much as possible.

5. Using your result from exercise 4, show how D changes as S gets large and n remains fixed. Varying S allows us to see the effect of increasing species richness.

   Find $\lim_{S \to \infty} \left(D(S, n) \right)$.

6. Using your result from exercise 4, show how D changes as n gets large but as S remains fixed. Varying n as we have done previously allows us to see the effect of increasing the numbers of individuals when the population proportions stay constant.

   Find $\lim_{n \to \infty} \left(D(S, n) \right)$.