Gambling with Genetics: An Introduction to Probability and Chi Square Analysis

Susan A. Moore, PhD
Duquesne University
Department of Biological Sciences
246 Mellon Building
600 Forbes Ave
Pittsburgh, PA 15282

Biography

Susan A. Moore obtained her B.S. from Pennsylvania State University and her M.S. and Ph.D. from the University of Michigan. She is currently an Assistant Professor in the Biological Science Department at Duquesne University, where she teaches a General Biology class for majors, an introductory biology course for non-majors, genetics, and immunology. Her research concerns teaching and learning in the large classroom.

Introduction

Probability can be defined as the number of times an event occurs divided by the total of opportunities for it to occur. Genetics is really a matter of probability, the likelihood of the occurrence of a particular outcome. Mendel understood probability and proposed the transmission of traits from one generation to another could be predicted by the laws of chance.

To take a simple example, consider the probability of coming up with heads in a single toss of a coin is one chance in two, or ½, and the likelihood that it comes up tails is also ½. A coin flip has only two possible outcomes heads or tails and the probability that it will be either heads or tails is ½ + ½ = 1. The sum of all probabilities of all possible outcomes is 1. This is the **Additive or Sum Rule of probability**.

To continue with the example of the coin toss, the probability of two events occurring simultaneously, or tossing the coin twice and getting heads twice, is the product of the two individual probabilities, or ½ x ½ = ¼. The probability for getting heads on the first flip is ½ and the probability for getting heads on the second toss is ½ and the probability for both events is ¼. This is the **Multiplicative or Product Rule of Probability** and it is commonly used in genetics to develop hypotheses.

In order to determine whether the hypothesis is supported or falsified by the data obtained in an experiment, geneticists typically use the **chi-square** statistical test. In short, this test expresses the difference between expected (hypothetical) and observed (collected) numbers as a single value, chi square (chi2). If the difference between observed and expected results is large, a large chi2 results and it is time to seek a new hypothesis to explain the data, while a small difference results in a small chi2—hypothesis is supported. Chi2 values are calculated according to the formula:

$$(\text{Observed Value} - \text{Expected Value})^2 \div (\text{Expected Value})$$

In addition, we must establish degrees of freedom. The degrees of freedom represent the number of ways the observed categories or phenotypes can vary. It is always 1 less than the number of categories (n) of possible outcomes ($n - 1$).

Student Outline

Exercise 1: Gambling with Chi Square

The first activity moves away from the simple coin toss with only two possible outcomes, to using dice that has six possible outcomes. You will use the Additive and Multiplicative rules of probability to predict the different outcomes of rolling dice.

1. Calculate the probability of rolling a 3 on one dice. Write the probability as a hypothesis.

2. Using the Additive or Sum Rule of probability determine the probability of rolling two dice and getting a three on either dice. Write the probability as a hypothesis.

3. Using the Multiplicative or Product Rule of probability determine the probability of rolling two dice and getting ones on both dice (snake eyes). Write the probability as a hypothesis

4. Test each of the above hypotheses. When playing craps in a casino, the shooter is required to bounce the dice off the wall, to ensure that the dice rotates on all axes, to prevent "cheating". You will need to do the same using the "lab-fashioned crap table." Roll both dice 60 times and record the results of the outcome in the Table 1.

Table 1. Rolling the dice.

	1's	2's	3's	4's	5's	6's	Totals
a. Probability of rolling a 3 on one dice							1
Observed Black dice							
Observed Red dice							
	Rolled a 3			no 3 rolled			Totals
b. Probability of rolling a 3 when rolling two dice							1
Observed							
	Snake eyes			No Snake eyes			Totals
c. Probability of rolling two 1's when rolling two dice							1
Observed							

Suggestion: Record, in Table 1, the outcome for all 3 tests by rolling 2 dice (you have been provided with 2 different colored dice for this). So as you are testing the rolling of a 3 with 1 dice (Table 1: probability a), you can use the same rolls to test for rolling a 3 with two dice (Table 1: probability b) and for rolling snake eyes (Table 1: probability c).

5. For each problem, determine the "expected" by multiplying the probability of the specific outcome by the total rolls of dice, determine the degrees of freedom, calculate chi square, and record in Table 2.

Note: We use the "raw numbers" of the data and compare it to the "raw numbers" we expected. We do not use probability or ratios or fractions in chi2.

What does the Chi2 value mean? Do you need a new hypothesis?

Now, how do we interpret the chi2 value we found? Suppose the expected and observed values were identical, then the chi2 value is zero. You might guess that a number very close to zero indicates close agreement between observed and expected and a large chi2 value suggests that "something unusual" is taking place. We can "*eyeball"* the differences between the expected and observed and describe them as "important." But it is better to determine if the differences are statistically significant. Scientists use the term "statistical significance" to described observed differences considerably different from zero that could not occur by chance. The problem is that chance alone almost always causes small deviations between observed and expected results, even when the hypothesis being tested is correct.

When does the chi2 value indicate that chance alone cannot explain the deviation? Geneticists generally agree on a probability value of 1 in 20 (or 5% = 0.05) as the lowest acceptable value derived from the chi2 test. This number indicates that if the experiment is repeated many times, the deviations expected due to chance alone will be as large as or larger than those observed only about 5% or less of the time. Probabilities equal to or greater than 0.05 are considered to support the hypothesis, while probabilities lower than 0.05 do not support the hypothesis. Here we must consult a table of chi2 values to make our decision. Most genetics textbook contain a table of chi2 values.

6. From your chi2 values and the degrees of freedom, determine if the differences of your data are due to chance and your hypothesis is supported (probability values greater than 0.05) or it is time to get a new hypothesis (probability values less than 0.05).

7. The Chi-square test is a kind of "mathematical judge" of probabilities. If you were in a casino, gambling, would you accept the 5% limit of chance? If the dice just didn't roll the way you expected, would you suspect something else? Loaded dice, perhaps? How would you determine if your dice are "fair?" Are **YOUR** dice fair?

8. Write a short summary for each situation tested. Is your hypothesis supported or rejected? Does the data "fit good" with the hypothesis or are they significantly different that it requires the hypothesis to be rejected? If rejected, offer an alternative hypothesis.

Table 2. Rolling the dice Chi square analysis

	1's	2's	3's	4's	5's	6's	Totals
a. Probability of rolling a 3 on one dice							1
Expected (same for Black dice and for Red dice)							
Black dice Observed							
Black dice $(Obs - Exp)^2 \div (Exp)$							χ^2 df
Red dice Observed							
Red dice $(Obs - Exp)^2 \div (Exp)$							χ^2 df
	3 present			no 3 present			Totals
b. Probability of rolling a 3 when rolling two dice							1

Expected			
Observed			
$(Obs - Exp)^2 \div (Exp)$			χ^2 df
	Snake eyes	No Snake eyes	Totals
c. Probability of rolling two 1's when rolling two dice			1
Expected			
Observed			
$(Obs - Exp)^2 \div (Exp)$			χ^2 df

Exercise 2: Let the chips fall where they may

The second activity applies what we learned from the dice to genetics. In place of dice with six possible outcomes, we look at genes that in a diploid organism have two outcomes. We are using poker chips and from this point on we will refer to them as "gene chips." "Gene chips" are used much as coins have been used in the past, as representations of the genotypes of an individual. Flipping the "gene chips" represent the process of meiosis and the formation of gametes. This also represents the Mendelian Laws of Segregation and Independent Assortment. You will calculate the probability of specific offspring using the Additive and Multiplicative rules of probability learned in the first exercise and you will also use Punnett square, a simple tool for predicting offspring, to confirm or to verify your calculations.

1. You have been given two plastic bags of chips. Each bag has a gender chip with pink representing an X chromosome and blue, the Y chromosome. So, if your gender chip has pink on both sides, it is female. If it is pink on one side and blue on the other, it is male. In addition to gender, these individuals contain 3 other genes: the "A" (white), "B" (red), and "C" (blue) genes. Examine your gene chips of your "plastic" people. Each side of your gene chip represents one allele.
Determine the genotype of your plastic people.

2. For each individual, determine all of the possible gametes they can form from meiosis. (Ignore gender to keep it simple) One way to do this is to "tree". For example, if your genotype is AaBbCC then the tree would look like this:

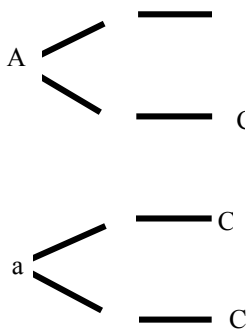

Figure 1: Determination of possible gametes by constructing a tree of possibilities. Here the possible gametes would be: ABC, AbC, aBC, and abC.

3. Let's keep things simple and ignore gender. Using the Multiplicative and the Sum Rules of Probability, calculate the probability of the following offspring from the cross:

parent 1 x parent 2
AaBbCc x AaBbCc

a. Offspring: AA bb CC

b. Offspring: aa BB cc

c. Offspring: Aa Bb Cc

Example: Probability of offspring with the genotype of Aa BB CC

P(Aa) x P(BB) x P(CC) Note use of Multiplicative rule
$P(Aa) = P(A_1a_2) + P(A_2a_1) = (½ \times ½) + (½ \times ½) = ½$ Note use of Additive rule
$P(BB) = ½ \times ½ = ¼$
$P(CC) = ½ \times ½ = ¼$
$P(AaBBCC) = ½ \times ¼ \times ¼ = 1/32$

4. Set up a Punnett square with the gametes listed for #2.

5. Determine from the Punnett square, how many offspring have the genotypes: AaBBCC, AAbbCC, aaBBcc, AaBbCc. **Are your calculated probabilities from #3 the same as what is determined by the Punnett square?**

6. Test the hypothesis of getting an offspring with the genotype of AaBbCc, from a cross with parents who are AaBbCc x AaBbCc. Much as a coin flip, tossing the gene chips represents meiosis. One person will flip the male gene chips and at the same time, another will flip the female's gene chips. The combined results represent the genotype of the offspring. Record data of 50 offspring and indicate how many have the AaBbCc genotype and how many do not.

7. Analyze your data and determine if your data supports your hypothesis (calculated probability).

Table 3 Gene chips Chi square analysis.

	AaBbCc Offspring	Other Offspring	Total
Probability			1
Observed			
Expected			
$(Obs - Exp)^2 \div (Exp)$			χ^2 df

8. Write a short summary. Is your hypothesis supported or rejected? Does the data "fit good" with the hypothesis or are they significantly different that it requires the hypothesis to be rejected? If rejected, offer an alternative hypothesis.

Notes for Instructor

Many students come into the sophomore level genetics class with little of no understanding of statistics and data analysis and these are important in the study of genetics. However, statistics can be uninteresting and tedious to teach and to learn. Presenting probability and chi square analysis in a context of gambling is a fun way to keep students engaged in the genetics classroom. These exercises connect scientific method, data analysis, mathematics and biology together, and are flexible enough for the introductory general biology student and for the genetics student. By decreasing or increasing the number of chips or traits, students can investigate monohybrid, dihybrid, or even trihybrid crosses. By adding "alleles" to the gender chips, the concept of sex-linked traits can be covered. These exercises can be done in the classroom or incorporated into a lab.

Materials

Weighted and normal or "fair" dice are needed for this exercise. Fair dice can be altered so that they no longer roll fair by filling in the pips of the dice or by cooking dice (plastic) in the microwave. See the following link for more information: http://homepage.ntlworld.com/dice-play/DiceCrooked.htm or http://en.wikipedia.org/wiki/Dice#Crooked_dice. Weighted dice can be obtained from Madhatter Magic Shop http://www.madhattermagicshop.com/magicshop/. In addition, to give the lab a "Los Vegas" flair, a lab-fashioned craps table is constructed from green felt and a vertical surface (wall, step, or even a cardboard wall). The green felt has an added benefit of keeping the dice from bouncing off the table.

For the second exercise, "gender chips" are made by gluing pink and blue poker chips together with pink and pink combination representing female and pink and blue representing male. Other color poker chips can be used to represent different traits and each face of the poker chip can

be marked to represent different alleles. Letters representing alleles are written on the surface of the poker chip using an marker.