Probabilities in genetics

The sum rule and product rule. Applying these rules to solve genetics problems involving many genes.

Introduction

The punnet square is a valuable tool, but it's not ideal for every genetics problem. For instance, suppose you were asked to calculate the frequency of the recessive class not for an Aax Aa cross, not for an AaBb x AaBb cross, but for an AaBbCcDdEe x AaBbCcDdEe cross. If you wanted to solve that question using a punnet square, you could do it – but you'd need to complete a punnet square with 1024 boxes. Probably not what you want to draw during an exam, or any other time, if you can help it!

The five-gene problem above becomes less intimidating once you realize that a punnet square is just a visual way of representing probability calculations. Although it's a great tool when you're working with one or two genes, it can become slow and cumbersome as the number goes up. At some point, it becomes quicker (and less error-prone) to simply do the probability calculations by themselves, without the visual representation of a clunky punnet square. In all cases, the calculations and the square Probabilities in genetics provide the same information, but by having both tools in your belt, you can be prepared to handle a wider range of problems in a more efficient way.

In this article, we'll review some probability basics, including how to calculate the probability of two independent events both occurring (event X and event Y) or the probability of either of two mutually exclusive events occurring (event X or event Y). We'll then see how these calculations can be applied to genetics problems, and, in particular, how they can help you solve problems involving relatively large numbers of genes.

Probability basics

Probabilities are mathematical measures of likelihood.

In other words, they're a way of quantifying (giving a specific, numerical value to) how likely something is to happen. A probability of 1 for an event means that it is guaranteed to happen, while a probability of 0for an event means that it is guaranteed not to happen. A simple example of probability is having a ½ chance of getting heads when you flip a coin.

Probabilities can be either empirical, meaning that they are calculated from real-life observations, or theoretical, meaning that they are predicted using a set of rules or assumptions.

The empirical probability of an event is calculated by counting the number of times that event occurs and dividing it by the total number of times that event could have occurred. For instance, if the event you were looking for was a wrinkled pea seed, and you saw it 1,850times out of the 7,324 Solution to the five-gene cross problem total seeds you examined, the empirical probability of getting a wrinkled seed would be 1,850/7,324 = 0.253, or very close to 1in 4 seeds.

The theoretical probability of an event is calculated based on information about the rules and circumstances that produce the event. It reflects the number of times an event is expected to occur relative to the number of times it could possibly occur . For instance, if you had a pea plant Heterozygous for a seed shape gene (Rr) and let it self-fertilize, you could use the rules of probability and your knowledge of genetics to predict that 1 out of every 4offspring would get two recessive alleles (rr) and appear wrinkled, corresponding to a 0.25(1/4) probability . We'll talk more below about how to apply the rules of probability in this case.

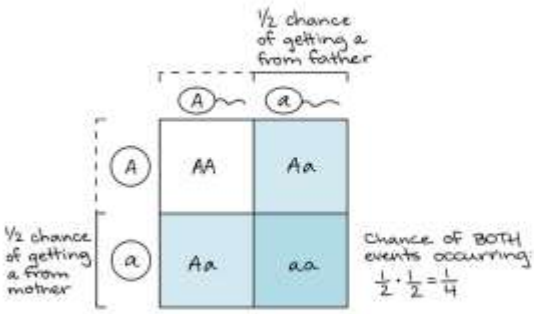
In general, the larger the number of data points that are used to calculate an empirical probability, such as shapes of individual pea seeds, the more closely it will approach the theoretical probability.

The product rule

One probability rule that's very useful in genetics is the product rule, which states that the probability of two (or more) independent events occurring together can be calculated by multiplying the individual probabilities of the events. For example, if you roll a six-sided die once, you have a 1/6chance of getting a six. If you roll two dice at once, your chance of getting two sixes is: (probability of a six on die 1) x (probability of a six on die 2) = (1/6) · (1/6) = 1/36.

In general, you can think of the product rule as the "and" rule: if both event X and event Y must happen in order for a certain outcome to occur, and if X and Y are independent of each other (don't affect each other's likelihood), then you can use the product rule to calculate the probability of the outcome by multiplying the probabilities of X and Y.

We can use the product rule to predict frequencies of fertilization events. For instance, consider a cross between two heterozygous (Aa) individuals. What are the odds of getting an aa individual in



the next generation? The only way to get an aa individual is if the mother contributes an a gamete and the father contributes an a gamete. Each parent has a ½ chance of making an a gamete. Thus, the chance of an aa offspring is: (probability of mother contributing a) x (probability of father contributing a) = (1/2) · (1/2) = 1/4.

This is the same result you'd get with a punnet square, and actually the same logical process as well

—something that took me years to realize! The only difference is that, in the punnet square, we'd do the calculation visually: we'd represent the 1/2probability of an a gamete from each parent as one out of two columns (for the father) and one out of two rows (for the mother). The 1-square intersect of the column and row (out of the 4total squares of the table) represents the 1/4chance of getting an a from both parents.

The sum rule of probability

In some genetics problems, you may need to calculate the probability that any one of several events will occur. In this case, you'll need to apply another rule of probability, the sum rule. According to the sum rule, the probability that any of several mutually exclusive events will occur is equal to the sum of the events' individual probabilities.

For example, if you roll a six-sided die, you have a 1/6 chance of getting any given number, but you can only get one number per roll. You could never get both a one and a six at the same time; these outcomes are mutually exclusive. Thus, the chances of getting either a one or a six are: (probability of getting a 1) + (probability of getting a 6) = (1/6) + (1/6) = 1/3.

You can think of the sum rule as the "or" rule: if an outcome requires that either event X or event Y occur, and if X and Y are mutually exclusive (if only one or the other can occur in a given case), then the probability of the outcome can be calculated by adding the probabilities of X and Y .

As an example, let's use the sum rule to predict the fraction of offspring from an Aax Aa cross that will have the dominant phenotype (AA or Aa genotype). In this cross, there are three events that can lead to a dominant phenotype:

Two A gametes meet (giving AA genotype), or

A gamete from Mom meets a gamete from Dad

(giving Aa genotype), or

A gamete from Mom meets A gamete from Dad

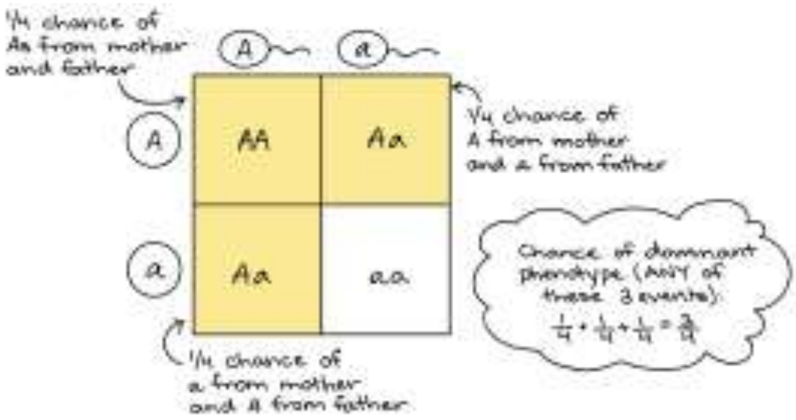
(giving Aa genotype)

In any one fertilization event, only one of these three possibilities can occur (they are mutually exclusive). Since this is an "or" situation where the events are mutually exclusive, we can apply the sum rule. Using the product rule as we did above, we can find that each individual event has a probability of 1/4. So, the probability of offspring with a dominant phenotype is:

(probability of A from Mom and A from Dad) +

(probability of A from Mom and a from Dad) +

(probability of a from Mom and A from Dad) = (1/4) + (1/4) + (1/4) = 3/4.



Once again, this is the same result we'd get with a punnet square. One out of the four boxes of the punnet square holds the dominant homozygote, AA. Two more boxes represent heterozygotes, one with a maternal A and a paternal a, the other with the opposite combination. Each box is 1 out of the 4 boxes in the whole punnet square, and since the boxes don't overlap (they're mutually exclusive), we can add them up (1/4 + 1/4 + 1/4 = 3/4) to get the probability of offspring with the dominant phenotype.

The product rule and the sum rule

Product rule Sum rule

Applying probability rules to dihybrid crosses

Direct calculation of probabilities doesn't have much advantage over punnet squares for single-gene inheritance scenarios. (In fact, if you prefer to learn visually, you may find direct calculation trickier rather than easier.) Where probabilities shine, though, is when you're looking at the behaviour of two, or even more, genes.

For instance, let' s imagine that we breed two dogs with the genotype BbCc, where dominant allele B specifies black coat colour (versus b, yellow coat colour) and dominant allele C specifies straight fur (versus c, curly fur). Assuming that the two genes assort independently and are not sex-linked, how can we predict the number of BbCc puppies among the offspring?

One approach is to draw a 16-square punnet square. For a cross involving two genes, a punnet square is still a good strategy. Alternatively, we can use a shortcut technique involving four-square punnet squares and a little application of the product rule. In this technique, we break the overall question down into two smaller questions, each relating to a different genetic event:

1. What’s the probability of getting a Bb genotype?

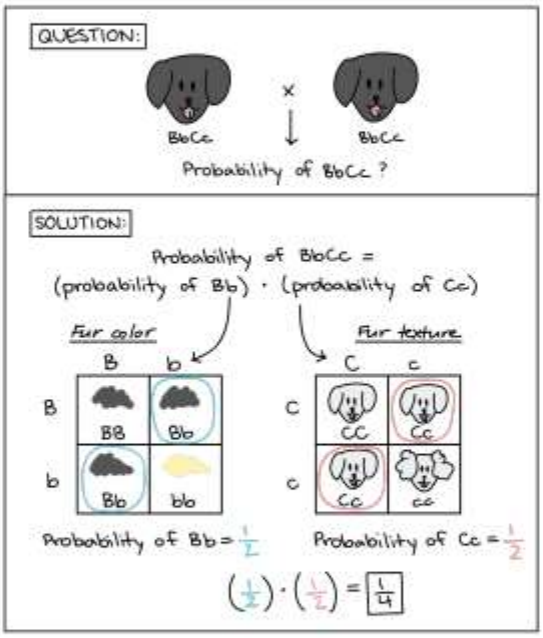
2. What’s the probability of getting a Cc genotype?

For independent events X and Y, the probability (P) of them both occurring (X and Y) is P (X) · P (Y).

For mutually exclusive events X and Y, the probability (P) that one will occur (X or Y) is ( ) + ( )

In order for a puppy to have a BbCc genotype, both of these events must take place: the puppy must receive Bb alleles, and it must receive Cc alleles. The two events are independent because the genes assort independently (don't affect one another's inheritance).

So, once we calculate the probability of each genetic event, we can multiply these probabilities using the product rule to get the probability of the genotype of interest (BbCc).



To calculate the probability of getting a Bb genotype, we can draw a 4-square punnet square using the parents' alleles for the coat colour gene only , as shown above. Using the punnet square, you can see that the probability of the Bb genotype is 1/2. (Alternatively, we could have calculated the probability of Bb using the product rule for gamete contributions from the two parents and the sum rule for the two gamete combinations that give Bb.) Using a similar punnet square for the parents' fur texture alleles, the probability of getting a Cc genotype is also 1/2. To get the overall probability of the BbCc genotype, we can simply multiply the two probabilities, giving an overall probability of 1/4.