

# 015\_Rosalind\_LIA

June 25, 2019

## 1 Independent Alleles

### 1.1 Mendel's Second Law

Recall that Mendel's first law states that for any factor, an individual randomly assigns one of its two alleles to its offspring. Yet this law does not state anything regarding the relationship with which alleles for different factors will be inherited.

After recording the results of crossing thousands of pea plants for seven years, Mendel surmised that alleles for different factors are inherited with no dependence on each other. This statement has become his second law, also known as the law of independent assortment.

What does it mean for factors to be "assorted independently?" If we cross two organisms, then a shortened form of independent assortment states that if we look only at organisms having the same alleles for one factor, then the inheritance of another factor should not change.

For example, Mendel's first law states that if we cross two **Aa** organisms, then 1/4 of their offspring will be **aa**, 1/4 will be **AA**, and 1/2 will be **Aa**. Now, say that we cross plants that are both *heterozygous* for two factors, so that both of their genotypes may be written as **Aa Bb**. Next, examine only **Bb** offspring: Mendel's second law states that the same proportions of **AA**, **Aa**, and **aa** individuals will be observed in these offspring. The same fact holds for **BB** and **bb** offspring.

As a result, independence will allow us to say that the probability of an **aa BB** offspring is simply equal to the probability of an **aa** offspring times the probability of an **BB** organism, i.e.,  $1/4 * 1/4 = 1/16$ . Because of independence, we can extend the idea of *Punnett squares* to multiple factors. We now wish to quantify Mendel's notion of independence using probability.

### 1.2 Problem

If two events *A* and *B* are independent then **Pr(A and B)** is equal to **Pr(A) x Pr(B)**. In other words, the events do not influence each other, so that we may simply calculate each of the individual probabilities separately and then multiply.

More generally, random variables **X** and **Y** are independent if whenever *A* and *B* are respective events for *X* and *Y*, *A* and *B* are independent (i.e., **Pr(A and B) = Pr(A) x Pr(B)**).

As an example of how helpful independence can be for calculating probabilities, let **X** and **Y** represent the numbers showing on two six-sided dice. Intuitively, the number of pips showing on one die should not affect the number showing on the other die. If we want to find the probability that **X + Y** is odd, then we don't need to draw a tree diagram and consider all possibilities. We simply first note that for **X + Y** to be odd, either *X* is even and *Y* is odd or *X* is odd and *Y* is even.

In terms of probability:

$$Pr(X + Y \text{ is odd}) = Pr(X \text{ is even and } Y \text{ is odd}) + Pr(X \text{ is odd and } Y \text{ is even})$$

Using independence, this becomes:

$$[Pr((X \text{ is even}) * Pr(Y \text{ is odd}))] + [Pr((X \text{ is odd}) * Pr(Y \text{ is even}))]$$

or

$$\left(\frac{1}{2}\right)^2 + \left(\frac{1}{2}\right)^2 = \frac{1}{2}$$

### 1.2.1 Given

Two positive integers  $k$  ( $k \leq 7$ ), the number of generations, and  $N$  ( $N \leq 2^k$ ), the number of Het offspring.

In this problem, we begin with Tom, who in the 0th generation has genotype Aa Bb. Tom has two children in the 1st generation, each of whom has two children, and so on. Each organism always mates with an organism having genotype Aa Bb.

### 1.3 Return

The probability that at least  $N$  Aa Bb organisms will belong to the  $k$ -th generation of Tom's family tree (don't count the Aa Bb mates at each level). Assume that Mendel's second law holds for the factors.

## 2 Solution - Bernoulli Trials

### 2.1 Probability of getting the double hetrozygous allele

If an organism mates with a organism which has a hetrozygous allele, no matter what the other organisms allele type is, it will have a 50% chance of its offspring being hetrozygous. For two alleles, if independent, the probability of two alleles being hetrozygous is the product of the probability of getting each hetrozygous allele.

$$Pr(AaBb) = Pr(Ab) * Pr(Bb)$$

$$Pr(AaBb) = \frac{1}{2} * \frac{1}{2} = \frac{1}{4}$$

### 2.2 Bernoulli

This is a case of an offspring having both hetrozygous alleles or not. So it can be modelled by a Bernoulli Trial (success or failure). For independent trials, with two outcomes (double hetrozygous or not), the probability of  $N$  successes in  $n$  trials is denoted as:

$$P(N \text{ successes in } n \text{ trials}) = \binom{n}{N} p^N * q^{n-N}$$

where  $p$  is the probability that it will occur (in this case 0.25) and  $q$  is the probability it won't (0.75).

## 2.3 Population

For each offspring there will be a separate probability. To get the final result we need to sum independent probabilities. The total number of offspring,  $O$ , will be  $k^2$ .

$$result = \sum_{i=N}^O \binom{O}{i} 0.25^i * 0.75^{O-i}$$

```
[2]: import math
```

```
[7]: # function to calculate probability of the f
def berHet(k, N):
    result = 0
    offspring = 2**k

    # loop from the number of successes (double Het offspring) to the
    ↪population
    for i in range(N, offspring + 1):
        prob = (math.factorial(offspring) / (math.factorial(i) * math.
    ↪factorial(offspring - i))) * (0.25**i) * (0.75**(offspring - i))
        result += prob

    print(result)
```

```
[5]: # Test data
berHet(2,1)
```

0.68359375

```
[6]: # actual data
berHet(5,9)
```

0.40648834849733056