Introduction to Mendelian Inheritance

Jan Emmanuel Samson

2024-07-29

Problem

Probability is the mathematical study of randomly occurring phenomena. We will model such a phenomenon with a random variable, which is simply a variable that can take a number of different distinct outcomes depending on the result of an underlying random process.

For example, say that we have a bag containing 3 red balls and 2 blue balls. If we let X represent the random variable corresponding to the color of a drawn ball, then the probability of each of the two outcomes is given by $\Pr(X = \text{red}) = \frac{3}{5}$ and $\Pr(X = \text{blue}) = \frac{2}{5}$.

Random variables can be combined to yield new random variables. Returning to the ball example, let Y model the color of a second ball drawn from the bag (without replacing the first ball). The probability of Y being red depends on whether the first ball was red or blue. To represent all outcomes of X and Y, we therefore use a probability tree diagram. This branching diagram represents all possible individual probabilities for X and Y, with outcomes at the endpoints ("leaves") of the tree. The probability of any outcome is given by the product of probabilities along the path from the beginning of the tree.

An event is simply a collection of outcomes. Because outcomes are so distinct, the probability of an event can be written as the sum of the probabilities of its constituent outcomes. For our colored ball example, let A be the event "Y is blue". $\Pr(A)$ is equal to the sum of probabilities of two different outcomes:

$$\Pr(X = \text{blue and } Y = \text{blue}) + \Pr(X = \text{red and } Y = \text{blue})$$

$$= \frac{3}{10} + \frac{1}{10} = \frac{2}{5}$$

- **Given**: Three positive integers k, m, and n, representing a population containing $k + m + n$ organisms: k individuals are homozygous dominant for a factor, m are heterozygous, and n are homozygous recessive.

- **Return**: The probability that two randomly selected mating organisms will produce an individual possessing a dominant allele (and thus displaying the dominant phenotype). Assume that any two organisms can mate.

Sample Dataset

```
2 2 2
```

Sample Output

```
0.78333
```

Intuition

The problem is asking for the probability of an offspring having at least one dominant allele. Since there are three possible genotypes, the solution can be expressed as the joint probability of getting either a homozygous dominant (HH) or heterzygous (Hh) offspring:

$$\Pr(\text{offspring} = HH \mid \text{offspring} = Hh)$$

Alternatively, the complement of the probability of getting homozygous recessive offspring would be equivalent to the equation above. As this would require fewer computations, we will express the formula for the solution as:

$$1 - \Pr(\text{offspring} = hh)$$

We proceed by computing the probabilities of all possible parent combinations capable of producing a homozygous dominant offspring:

Mother	Father	Probability
Hh	Hh	$\frac{m}{t} * \frac{m-1}{t-1} * \frac{1}{4}$
Hh	hh	$\frac{m}{t} * \frac{n}{t-1} * \frac{1}{2}$
hh	Hh	$\frac{n}{t} * \frac{m}{t-1} * \frac{1}{2}$
hh	hh	$\frac{n}{t} * \frac{n-1}{t-1}$

We then sum all the probabilities and get the complement to get the final answer:

$$1 - \frac{0.25(m^2 - m) + mn + (n^2 - n)}{t^2 - t}$$

Solution

```python
def compute_p_recessive_complement(k: int, m: int, n: int) -> float:
    t = (k + m + n)
    p_mm = (m/t) * ((m-1)/(t-1)*0.25)        # mother = m, father = m
    p_mn = (m/t) * (n/(t-1)*0.5)             # mother = m, father = n
    p_nm = (n/t) * (m/(t-1) * 0.5)           # mother = n, father = m
    p_nn = (n/t) * ((n-1)/(t-1))             # mother = n, father = n
    p_homozygous = p_mm + p_mn + p_nm + p_nn # probability of getting homozygous child
    return round(1 - p_homozygous, 6)
```

```python
k, m, n = 2, 2, 2,
result = compute_p_recessive_complement(k, m, n)
print(result)
```

```
0.783333
```

Bibliography