Mathematics in Health Sciences

Spring 2020 Notes

Alexander Kasiukov

May 11, 2020

Contents

This is the reading and homework for the the week starting March 30, 2020. Please send me your solutions for the homework problems before midnight Friday, April 3. After that deadline, you will need to ask for new problems to make up for this homework.

1 Combinatoral Properties of Genetic Inheritance

These notes intentionally overlook some of the less significant details of genetics and genetic inheritance. I will also use mathematical language in a way which is not usually seen in introductory biology texts.

1.1 Inheritance of Traits

Individual members of the same species may have distinct differences in some of their biological characteristics. For example, people may have different eye color. These characteristics are called `traits`. Traits are inherited, in the sense that a particular trait of the child is probabilistically dependent on the same trait of the parent.

Gregor Mendel was perhaps the first person who came up with the idea of an atomic inheritable component of a trait, around 1860. That idea found its realization in the modern concept of `gene`.

In the mid-20th century, the chemical compound called DNA was discovered inside of living organisms. Rosalind Franklin and Maurice Wilkins studied the structure of DNA using X-ray crystallography. Their work led James D. Watson and Francis Crick to the hypothesis that DNA was the carrier of inherited information. This hypothesis was confirmed, leading to the modern idea of `gene` as a particular place in the DNA of the organism.

1.2 Structure of DNA

DNA (Deoxyribonucleic acid) is a long organic polymer molecule, composed of two polynucleotide chains connected to each other, similar in struc-

ture to rail track (with the polynucleotide chains being analogous to the metal rails, and the connections between them being similar to the railroad ties). Unlike a straight railroad track, the two polynocleotide chains of a DNA molecule are twisted around each other, giving the DNA its famous "double helix" geometry.

The two polynucleotides chains of DNA are composed of simpler units called `nucleotides`. Each nucleotide is composed of a nitrogen-containing `nucleobase`, a sugar called deoxyribose, and a phosphate group. There are four possible nucleobases:

- cytosine [C],
- guanine [G],
- adenine [A],
- thymine [T][1].

The nucleotides are joined to one another in a chain by covalent bonds. (Using the rail track analogy, the covalent bonds produce the "rail" of the rail track.) The nucleobases of the two separate polynucleotide strands are bound together with hydrogen bonds. (The hydrogen bonds form the "ties" of the railroad track.) The pair of two nucleotides joined by hydrogen bond is called a `base pair`.

Adenine forms the bond only with thymine, and cytosine — only with guanine. (These are known as the `base pairing rules`.) Because of the base pairing rules, there are only two types of base pairs:

- adenine — thymine, and
- cytosine — guanine.

If one strand of DNA is known, the base pairing rules determine the other complimentary strand. This is the reason the DNA sequence is usually given as a sequence of nucleotides on one strand only.

Any three consecutive base pairs along the DNA strand are called a `triplet`[2] The ATG (and some other triplets) are called `start triplets` and TAA (as well as some other triplets) are called `stop triplets`.

[1] This nucleobase is replaced by uracil [U] in a slightly different molecule called `RNA`.

[2] The same concept for RNA is called a `codon`.

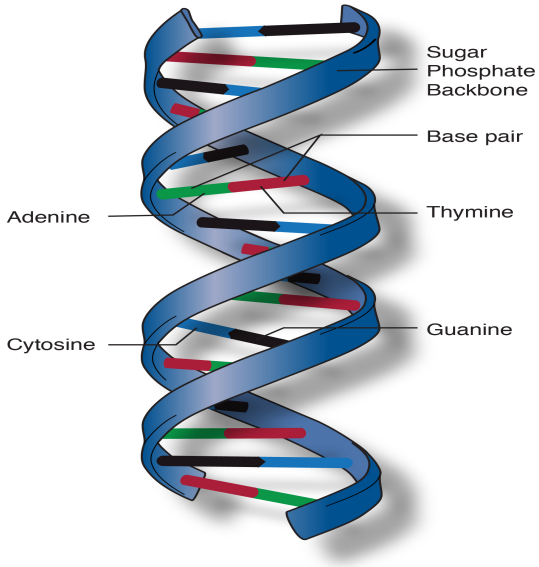

Figure 1: The DNA Double Helix[3]

1.3 Biological Role of DNA

In an organism, the DNA is arranged in `chromosomes`. Each chromosome is a single molecule of DNA with some `packaging proteins` around them. Chromosomes are inside of the cells of an organism. Within a cell, some chromosomes are inside of the cell's nucleus; those are called the `nuclear DNA`. Additionally, each mitochondrion [3] contains a single small circular chromosome. These chromosomes are called `mitochondrial DNA`.

Each species has a fixed number of nuclear chromosomes, classified in several types depending on their structure. Unlike nuclear chromosomes, all mitochondrial chromosomes are the same for a given species. Humans have 46, dogs — 78, and cats — 38 nuclear chromosomes in most[4] of their cells.

The chromosomes are used by the cell in production of proteins that make

[3]Mitochondria are cellular organelles within eukaryotic cells, and each human cell has between 1 and 2 thousands of them.

[4]Some cells, for example the erythrocytes (the red blood cells), contain no nucleus, and thus no nuclear DNA. Furthermore, some of the cells, considered in more detail in the next section, contain only half as many chromosomes in their nucleus as most other cells in the body.

up organism's body. Proteins are made up of amino acids, and each amino acid is encoded in the chromosome by a single triplet. Some triplets serve other functions as well, like the start and stop triplets. A specific distinct section of the chromosome between a start triplet and a stop triplet, which encodes the synthesis of a single protein, is called a `gene`.

The term 'gene' describes certain location on a chromosome of some specific type. The actual sequence of nucleotides at that location is called the `allele` of that particular gene. Thus, in this strict sense, all members of the same species have the same exact genes, but differ in their alleles.

The complete sequence of genes of a particular species is called its `genome`. Humans have around $25,000$ genes on the 24[5] types of their nuclear chromosomes. Those genes were identified, catalogued, and made publicly available by the Human Genome Project, completed in 2003. In contrast, human mitochondrial chromosome contains only 37 genes.

1.4 Diploid and Haploid Cells

Cells in human body are classified into two large groups: `somatic cells` that comprise most of the tissue, and `gametes` which are fundamental to sexual reproduction (thus they are also called reproductive or sex cells). Male gametes are called `sperm` and female — `ova` or egg cells.

The chromosomes that have the same structure are called `homologous`. In particular, homologous chromosomes have the same genes.

Somatic cells have 46 nuclear chromosomes that may be grouped in 22 homologous pairs [6], called `autosomal DNA` or `autosomes`, and the remaining pair, called `sex chromosomes`.

The types of autosomes are numbered 1 through 22 in a standard way. A somatic human cell contains two "chromosome 1" copies as the first homologous pair, two "chromosome 2" copies as the second homologous pair, ..., two "chromosome 22" copies — as the twenty second homologous pair.

The types of sex chromosomes in a somatic cell depend on sex of the organism. Females have two homologous chromosomes called `X-chromosome`. Males have one X-chromosome and one chromosome of different structure, called `Y-chromosome`. Both the X-chromosome and the Y-chromosome are

[5] We are counting distinct types of chromosomes in human genome: the 22 autosomes, the X-chromosome and the Y-chromosome. These are defined in the next section.

[6] Meaning that the two chromosomes comprising that pair have the same structure.

called sex chromosomes. This paired structure of chromosomes in a somatic cell is called `diploid`.

A gamete (a sperm or egg cell) contains 23 chromosomes, with only one chromosome of each kind being present. An egg cell always has one X-chromosome. A sperm cell may have either X or Y, depending on whether that sperm cell will produce a female or male offspring. This chromosomal structure of a gamete is called `haploid`.

1.5 Cell Division and Inheritance of Genes

We will now focus on the mechanism of inheritance that applies to human species. The mitochondrial DNA is passed unchanged from mother to her offspring. Inheritance of nuclear DNA is more complicated.

Somatic cells divide in two distinctly different ways. The first kind of cell division, called `mitosis`, creates two copies of the mother cell with the same exact genetic material duplicated in the two daughter cells (which are also somatic cells). Mitosis is necessary to replace existing cells, or to allow the organism to grow from a single cell called `zygote` from which it starts.

The second kind of cell division, called `meiosis`, is the process of producing two[7] gamets from a single somatic cell. Each daughter gamete contains only one half of the number of chromosomes of the original mother somatic cell. (More specifically, it has one chromosome for each homologous pair of the mother cell.) In meiosis, the original chromosomes of the mother cell are divided into genes and mixed up within homologous pairs, resulting in two new chromosomes with allele combinations not present in any single chromosome of the mother cell. Each daughter cell gets only one such new chromosome. This breaking and reassembly process is called `genetic recombination`. Thus each meiotic cell division produces unique combination of alleles in the daughter cells.

In female meiosis, the two X-chromosomes recombine just like the autosomal chromosomes do. In male meiosis, the two sex chromosomes of different type, the X-chromosome and the Y-cromosome, do not recombine and are passed in their original form. One daughter cell gets the X-chromosome, and

[7]This is an oversimplification — the actual number is four — but it will help to focus on the significant part if we make this simplifying assumption.

the other — the Y-chromosome.

HOMEWORK: Cystic fibrosis is caused by mutations in a gene on chromosome 7 encoding a protein called the cystic fibrosis transmembrane conductance regulator. This gene is called CFTR gene.

1. How many alleles of CFTR gene does each human somatic cell have? (Solution: The CFTR gene is on chromosome 7 which is autosomal. Each somatic human cell has two autosomes of any specific type, in particular two chromosomes 7. Each of those two chromosomes has one allele of CFTR gene. Therefore a somatic human cell has two alleles of CFTR gene.)

2. How many alleles of CFTR gene does each human sperm cell have? (Solution: The CFTR gene is on chromosome 7 which is autosomal. Each gamete has one autosome of any specific type, in particular one chromosome 7. Therefore a sperm cell has one allele of CFTR gene.)

3. How many alleles of CFTR gene does each human ova cell have? (Solution: The CFTR gene is on chromosome 7 which is autosomal. Each gamete has one autosome of any specific type, in particular one chromosome 7. Therefore an ova cell has one allele of CFTR gene.)

HOMEWORK: Hemophilia A is caused by mutations in a gene on X-chromosome encoding a protein called factor VIII. This gene is called F8 gene.

1. How many alleles of F8 gene does each human somatic cell have? (Solution: The F8 gene is on X-chromosome, which a sex chromosome. Each female somatic human cell has two X-chromosomes. Therefore a somatic human cell of a female has two alleles of F8 gene. On the other hand, each male somatic cell has only one X-chromosome, therefore only one allele of F8 gene.)

2. How many alleles of F8 gene does each human sperm cell have? (Solution: The F8 gene is on X-chromosome, which a sex chromosome. A sperm cell may have either X-chromosome or Y-chromosome as its sex chromosome. In the first case it has one allele of F8 gene, and in the second — none.)

3. How many alleles of F8 gene does each human ova cell have? (Solution: The F8 gene is on X-chromosome, which a sex chromosome. An ova cell has one X-chromosome as its sex chromosome. Therefore an ova cell has one allele of F8 gene.)

Finally, when two gametes, a sperm cell and an ova cell, fertilize — they produce a single cell called `zygote` from which the whole organizm eventually grows by mitotic cell division. The zygote contains 46 chromosomes, 23 from the father and 23 from the mother. The sex of the organism that grows from the zygote is determined by the type of the sex chromosome passed by the father. If it is a Y-chromosome, the zygote will be male, and if the father passed X-chromosome — female.

1.6 Mendel's Laws

Johann Gregor Mendel was an Augustinian friar and abbot of St. Thomas' Abbey in Brünn, then in Austro-Hungarian Empire. (That city is the present-day Brno in Czech Repubic.) Mendel was experimenting with the abbey's

pea plants (*Pisum sativum*), observing how the physical appearance of the parent plants correlated with that of their offspring. He presented his findings at the February 8 and March 8, 1865 meetings of the Brünn Natural History Society, and his presentation was published in the Society's proceedings [4] in 1866.

In his study, Mendel developed the concept of a trait as an observable hereditary physical characteristic of an individual organism, and postulated the existence of hereditary factors that determine traits. He then concluded that each trait is a function of two of those factors, each inherited from one of the two parents. Mendel also discovered that some of the factors may be expressed as a trait, and others, while present in the organism of the parent, may not affect physical characteristic at all. This phenomenon led him to concepts of `dominant allele` and `recessive allele`. In his specific experiments Mendel studied many characteristics of the plants (color of their flowers and seeds, flower arrangement, length of the axis etc.). One example showing the phenomenon of dominant and recessive traits was the color of flowers. There were two types of plants: those with violet–red flowers and others with white flowers. Mendel observed the possibility of violet-red plants to have white offspring, thus demonstrating that those violet-red plants had the white flower factors in them, that were not directly observable as traits but could be passed to offsprings. On the other hand, white plants could only have white-flowered offspring, demonstrating that white-flowered plants did not contain any violet-red factors in them.

We now realize that each factor is nothing other than a gene, and the discovery of Mendel can be restated in modern terms by saying that an organism has two alleles of each gene, and inherits one of these alleses from the father and another one from the mother. Mendel's Laws are a subset of laws of inheritance that apply to the situation when a trait is determined by a single gene on autosomal DNA. Such a trait is called a `single locus trait` or a Mendelian trait. In contrast, a trait controlled by two or more genes is called a `polygenic trait`.

The full combination of traits of a given organism is called the `phenotype`, and the full combination of their alleles is called the `genotype`.

The following is a reformulation of Mendel's Laws in terms of the modern understanding of inheritance mechanism.

1.6.1 Law of Dominance and Uniformity

The darkness of eye color in humans (a trait) is controlled by the gene called OCA2 on chromosome 15[8]. There are two alleles of that gene, one (we denote it M_+) resulting in melanin produced in high amounts and another (we denote it M_-) resulting in reduced amount or quality of melanin.

An organism is called `homozygous` for a particular gene if and only if the two alleles of that gene in the organism's genotype are identical. Otherwise the organism is called `heterozygous`. We say that an allele is associated with a particular variant of the trait if and only if a homozygous carrier of that allele (i.e. an organism with two copies of that allele) displays that specific variant of that trait. For example, the allele M_+ is associated with dark eye color (particular variant of the eye color trait) because someone with two M_+ alleles will have dark eyes; the allele M_- is associated with blue eye color because someone with two M_- alleles will have blue eye color.

For any given gene, an allele of that gene is called dominant with respect to another allele of that gene (which is then called recessive with respect to the first) if and only if an organism with one dominant and one recessive allele in their genotype will display the trait associated with the dominant allele, masking the effect of the recessive allele. For example, M_+ allele is dominant with respect to M_- allele, and M_- is recessive with respect to M_+, because an individual with one M_+ and one M_- allele will have dark eyes.

The law of dominance and uniformity postulates that when a homozygous carrier of a dominant allele is mated with a homozygous carrier of a recessive allele, all offspring in the first generation will display the variant of the trait associated with the dominant allele.

1.6.2 Law of Segregation of Alleles

The parent will pass to their offspring one and only one of the two alleles they have for each particular gene, with equal chance (hence equal to 50%). For example, if one parent is blue eyed and another one is heterozygous dark eyed, then there is a 50-50 chance their child will be blue eyed.

[8]Eye color is controlled by several genes, so this example is not the full description of the eye color phenomenon, but rather a simplistic model that works in most — but not all — situations.

1.6.3 Law of Independent Assortment

Different genes (responsible for different traits) segregate independently from each other. For example, suppose a particular individual is heterozygous both with respect to the eye color OCA2 gene and cystic fibrosis CFTR gene. Then knowing that this individual's child has blue eyes (meaning that the individual under consideration passed their M_- allede to the child) does not help in determining if the child will have cystic fibrosis.

We can derive the following probabilistic consequences from Mendel's Laws:

1. For any given gene, an organism has two alleles, with one allele from the father, and another one from the mother. The allele received from one parent is independend from the allele received from another parent.

2. For any given gene, an organism has equal, i.e. $\frac{1}{2}$, chance of passing either allele to their own offspring.

3. For any given parent, the allele received from that parent for one gene is independent from the allele received from the same parent for another gene.

HOMEWORK: There is a family with a blue-eyed father and a dark-eyed mother. They have a blue-eyed daughter. What is the probability that their next child will have dark eyes?

Justify your answer using the properties of probability and the laws of inheritance.

$$P\begin{pmatrix}\text{next}\\ \text{child}\\ \text{has dark}\\ \text{eyes}\end{pmatrix} = \boxed{\text{Probability of negation}} = 1 - P\begin{pmatrix}\text{next}\\ \text{child}\\ \text{has blue}\\ \text{eyes}\end{pmatrix} =$$

$$= \boxed{\text{Blue eyes is recessive trait}} = 1 - P\begin{pmatrix}\text{both}\\ \text{child's}\\ \text{alleles}\\ \text{are } M_-\end{pmatrix} =$$

$$= \boxed{\text{One allele from each parent}} = 1 - P\left(\begin{pmatrix}\text{mother}\\ \text{gave}\\ M_-\\ \text{allele}\end{pmatrix} \cap \begin{pmatrix}\text{father}\\ \text{gave}\\ M_-\\ \text{allele}\end{pmatrix}\right) =$$

$$= \boxed{\text{Product rule for probability}} =$$

$$= 1 - P\begin{pmatrix}\text{mother}\\ \text{gave}\\ M_-\\ \text{allele}\end{pmatrix} \cdot P\left(\begin{pmatrix}\text{father}\\ \text{gave}\\ M_-\\ \text{allele}\end{pmatrix} \middle| \begin{pmatrix}\text{mother}\\ \text{gave}\\ M_-\\ \text{allele}\end{pmatrix}\right) =$$

$$= \boxed{\text{Father's allele independent from that of the mother}} =$$

$$= 1 - P\begin{pmatrix}\text{mother}\\ \text{gave}\\ M_-\\ \text{allele}\end{pmatrix} \cdot P\begin{pmatrix}\text{father}\\ \text{gave}\\ M_-\\ \text{allele}\end{pmatrix} =$$

$$= \boxed{\text{Mother heterozygous, father homozygous}} =$$

$$= 1 - \frac{1}{2} \cdot 1 = 1 - \frac{1}{2} = \frac{1}{2}.$$

2 Statistical Experiment with Hypotheses

2.1 The Concept of a Hypotheses System

Suppose we have a statistical experiment. A system of events $H_0, H_1, \ldots,$ associated with that statistical experiment are called `hypotheses` if and only if these events are:

1. `pairwise mutually exclusive`, meaning that for any pair H_i and H_j of two different events from that system, we have that

$$P(H_i \cap H_j) = 0,$$

2. `collectively exhaustive`, meaning that the union of all these events

$$\bigcup_{i \in \{0,1,\ldots\}} H_i$$

equals the whole sample space of the experiment in question.

These two conditions are equivalent to saying that any outcome of the experiment results in one and only one of these events taking place.

Example: In a statistical experiment with finately many outcomes, the events made up from those outcomes are a system of hypotheses. (This example is not particularly interesting.) In the experiment when we roll a fair die and get a number from $\{1, \ldots, 6\}$ as an outcome, the six events $H_0 = \{1\}, H_1 = \{2\}, \ldots, H_5 = \{6\}$ are a system of hypotheses.

Example: In the experiment where we shuffle a deck of cards and select the one on the top, the four card suits: hearts, spades, clubs and diamonds — form a system of hypotheses.

2.2 Formula of Total Probability

Theorem (*Formula of Total Probability*). *For any statistical experiment with a system of hypotheses $H_0, H_1, \ldots$, the probability of any event E associated with this experiment can be computed as:*

$$P(E) = P(H_0) \cdot P(E|H_0) + P(H_1) \cdot P(E|H_1) + \ldots$$

$\diamondsuit$

Example (Use of the Total Probability Formula in medical testing):
Mammogram is an X-ray imaging of human breast, used for diagnosis or screening of breast cancer. Approximately 1 in 8 US women (about 12.4%) will develop invasive breast cancer over the course of her lifetime: $P(\text{breat cancer}) = 12.4\%$. The following parameters can be defined for any diagnostic procedure. They are listed below with their estimated values for mammogram, based on [5]

- `sensitivity` of the test (also called the "true positive" rate)

$$P\left(\begin{matrix}\text{positive}\\ \text{mammogram}\end{matrix}\middle|\begin{matrix}\text{breast}\\ \text{cancer}\end{matrix}\right) = 90\%$$

  (estimates range from 83 to 95 percent);

- `specificity` of the test (also called the "true negative" rate)

$$P\left(\begin{matrix}\text{negative}\\ \text{mammogram}\end{matrix}\middle|\begin{matrix}\text{no breast}\\ \text{cancer}\end{matrix}\right) = 95\%$$

  (estimates range from 90 to 98 percent);

- type I error probability (also called the "false positive" rate)

$$P\left(\begin{matrix}\text{positive}\\ \text{mammogram}\end{matrix}\middle|\begin{matrix}\text{no breast}\\ \text{cancer}\end{matrix}\right) = 1 - (\text{specificity}) = 5\%;$$

- type II error probability (also called the "false negative" rate)

$$P\left(\begin{matrix}\text{negative}\\ \text{mammogram}\end{matrix}\middle|\begin{matrix}\text{breast}\\ \text{cancer}\end{matrix}\right) = 1 - (\text{sensitivity}) = 10\%;$$

Sensitivity and specificity of a mammogram depend on age of the patient and type of breast tissue. The above are the overall rates that should be adjusted in any specific case to take into account the particulars of the patient being tested.

Let's use the formula of total probability to find the probability that a random US woman having a mammogram will get a positive mammogram result.

$$P\left(\begin{matrix}\text{positive}\\ \text{mammogram}\end{matrix}\right) = \boxed{\text{formula of total probability}} =$$

$$P\left(\begin{matrix}\text{breast}\\ \text{cancer}\end{matrix}\right) \cdot P\left(\begin{matrix}\text{positive}\\ \text{mammogram}\end{matrix}\,\middle|\,\begin{matrix}\text{breast}\\ \text{cancer}\end{matrix}\right)$$

$$+ P\left(\begin{matrix}\text{no breast}\\ \text{cancer}\end{matrix}\right) \cdot P\left(\begin{matrix}\text{positive}\\ \text{mammogram}\end{matrix}\,\middle|\,\begin{matrix}\text{no breast}\\ \text{cancer}\end{matrix}\right)$$

$$= (\text{prevalence}) \cdot (\text{sensitivity}) + \big(1 - (\text{prevalence})\big) \cdot (\text{false positive rate})$$

$$= 12.4\% \cdot 90\% + (1 - 12.4\%) \cdot 5\% = 15.5\%.$$

♢

HOMEWORK: According to [6], one of the serological tests approved in the US for SARS-CoV-2 testing is the RDT test by Cellex Inc. The sensitivity of this test is reported to be 93.8% and specificity — 95.6%. Assuming that 10% of the New York city population is infected with SARS-CoV-2, find the probability that a randomly selected New Yorker will have a postitive result of the RDT test.

2.3 Bayes' Formula

The subject of this secton was discovered by a Presbyterian minister Thomas Bayes (c. 1701 – 1761). His work containing this result was published posthumously by Richard Price (1723 – 1791) in 1763 as [1].

Suppose we have a family with one blue eyed and one dark eyed parent. If a blue eyed child is born in this family, then the dark eyed parent is heterozygous. On the other hand, if a dark eyed child is born to the same two parents, it is still possible for the dark eyed parent to be either heterozygous or homozygous. So, from the *binary logic point of view*, the dark eyed child gives no information about the genotype of the dark eyed parent.

When the dark eyed parent is heterozygous, about half of the children born to these parents will be blue eyed, and another half — dark eyed. If we observe these two parents having 50 dark eyed children[9], then we would conclude with almost full certainty that the dark eyed parent must be homozygous. This suggests that from the *probabilistic point of view*, even a single dark eyed child gives — however small and incremental — information about the dark eyed parent's genotype. Bayes' formula is a way to quantify that information.

Theorem (*Bayes's Formula*). *Suppose we have a statistical experiment with a system of hypotheses $H_0, H_1, \ldots$, and another event E associated with this experiment. Assume that H_* is one of the hypotheses $H_0, H_1, \ldots$ Then we have that*

$$P(H_*|E) = \frac{P(H_*) \cdot P(E|H_*)}{P(H_0) \cdot P(E|H_0) + P(H_1) \cdot P(E|H_1) + \ldots}$$

$\diamondsuit$

Remark (about *structure of Bayes' formula*): The denominator of the fraction is the formula of total probability for the event E, and the numerator is the term of the denominator, that corresponds to the hypothesis under consideration, namely the H_*. The $P(H_i)$ is called the `a priori probability`[10]

[9] However far-fetched, this number is not unreal: Mrs. and Mr. Feodor Vassilyev (b. 1707–c.1782) are recoreded as the parents who had the highest number of children. Their children, 69 in total, were born between 1725 and 1765 in Shuya, Russia.

[10] Literally "before probability", i.e. our belief in the truth of H_i before we observe the event E.

of the hypothesis H_i, and the $P(H_i|E)$ — the `a posteriori probability`[11] of H_i. Thus, the Bayes' formula expresses the a posteriori probability of a particular hypothesis in terms of a priori probabilities of all hypotheses in the system, and conditional probabilities of the observed event E under each of those hypotheses. This is the reason this formula is sometimes called *the rational way to change your mind* based on the observed evidence. $\diamondsuit$

[11] Literally "after probability", i.e. our belief in the truth of H_i after we observe the event E.

Example (using the Bayes' formula): Suppose that in a family with one blue eyed and one dark eyed parent, there is a dark eyed child. Assuming that the probability of a dark eyed person being homozygous is p, we will have that:

$$P\left(\begin{array}{c|c}\text{dark eyed parent is homo-zygous} & \text{child is dark eyed}\end{array}\right) = \boxed{\text{Bayes formula}} =$$

$$= \frac{P\left(\text{dark eyed parent is homo-zygous}\right) \cdot P\left(\begin{array}{c|c}\text{child is dark eyed} & \text{dark eyed parent is homo-zygous}\end{array}\right)}{P\left(\text{dark eyed parent is homo-zygous}\right) \cdot P\left(\begin{array}{c|c}\text{child is dark eyed} & \text{dark eyed parent is homo-zygous}\end{array}\right) + P\left(\text{dark eyed parent is hetero-zygous}\right) \cdot P\left(\begin{array}{c|c}\text{child is dark eyed} & \text{dark eyed parent is hetero-zygous}\end{array}\right)}$$

$$= \boxed{P\left(\text{dark eyed parent is homo-zygous}\right) = p} = \frac{p \cdot 1}{p \cdot 1 + \left(1-p\right) \cdot \frac{1}{2}} =$$

$$= \boxed{\text{multiply the numerator and denominator by 2}} = \frac{2p}{2p + \left(1-p\right)} = \frac{2p}{p+1}.$$

So, for example, if $p = \frac{1}{2}$ (meaning that the probability of a homozygous and heterozygous dark eyed person in the population the same), the above formula yields:

$$\frac{2p}{p+1} = \frac{1}{\left(\frac{3}{2}\right)} = \frac{2}{3}.$$

In this case, the information about dark eyed child increases by $\frac{1}{6}$ (from $\frac{1}{2} = \frac{3}{6}$ to $\frac{2}{3} = \frac{4}{6}$) the probability of the dark eyed parent being homozygous. $\diamondsuit$

HOMEWORK: In a family with one blue eyed and one dark eyed parent, there are two dark eyed children. Assuming that the probability of a homozygous and heterozygous dark eyed person in the population the same, compute the probability of the dark eyed parent being homozygous.

2.4 Population Genetics

Consider one specific allele of a particular gene in the species' genome. By definition, `allele frequency` in a particular population of that species is the probability of that allele appearing in that gene of a a random chromosome selected from a random member of that population.

Example: if we have a family with one blue and one dark eyed parent and a blue eyed child, then the allele frequency of the M_- allele among the three members of that family is $\frac{5}{6}$. Indeed, in this family the dark eyed parent must be heterozygous (i.e. has only one M_- allele) while the other parent and the child each have two copies of M_- allele. Therefore only one among their 6 alleles of the eye color gene is M_+ and the remaining 5 alleles are M_-.

2.4.1 Hardy-Weinberg Principle

Theorem (*Hardy-Weinberg Principle*). *Fix a particular species and a specific gene in that species' genome. If the probability of an allele of that gene being passed to the next generation of an isolated population depends only on the frequency x of that allele in that populaton*[12] *then*

1. *the allele frequency x stays the same throughout generations;*
2. *as we go from one generation to the next, the frequency of homozygous genotype with that allele approaches x^2, the frequency of the heterozygous genotype with one such allele approaches $2x(1-x)$, and the frequency of all genotypes not having this allele approaches $(1-x)^2$;*
3. *the probability distribution described in the previous item is called the `Hardy-Weinberg equilibrium`, meaning that once the population is in that state, it will stay in that state indefinitely.*

$\diamondsuit$

The Hardy-Weinberg Principle can be used to determine the allele frequency based on the phenotype freqencies in the populaiton.

[12] Usually this conditsion is worded by giving several other conditions, collectively equivalent to this one: mating is random, there is no reproductive advantage or disadvantage conferred on individual by this allele, ...

Example (Use of Hardy-Weinberg Principle to determine genotype frequencies): According to [2], about one 1 of every 6 Americans has blue eyes:

$$P\begin{pmatrix}\text{blue}\\ \text{eye}\\ \text{color}\end{pmatrix} = \frac{1}{6} \approx 17\%$$

Assuming that the American pupulation is in Hardy-Weinberg equilibrium with respect to OCA2 gene that controls eye color, we can conclude that the freqency of the blue eye color is the square of the allele frequency M_-:

$$P\begin{pmatrix}\text{blue}\\ \text{eye}\\ \text{color}\end{pmatrix} = \left(P\begin{pmatrix}\text{frequency}\\ \text{of } M_-\\ \text{allele}\end{pmatrix}\right)^2$$

If we denote $x = P\begin{pmatrix}\text{frequency}\\ \text{of } M_-\\ \text{allele}\end{pmatrix}$, we can write the above equation as $\frac{1}{6} = x^2$.

The solution of this equation is $x = \sqrt{\frac{1}{6}} \approx 0.408 \approx 41\%$. Knowing this allele frequency, we can use the Hardy-Weinberg principle to work can our way backwards finding all other genotype frequencies. This work can be aided with the visual tool called `Punnett square`:

Father / **Mother**	**Event:** father gave M_+ **Probability:** $1-x \approx 59\%$	**Event:** father gave M_- **Probability:** $x = \sqrt{\frac{1}{6}} \approx 41\%$
Event: mother gave M_+ **Probability:** $1-x$	**Genotype:** M_+M_+ **Trait:** dark eyes **Probability:** $(1-x)^2 \approx 35\%$	**Genotype:** M_+M_- **Trait:** dark eyes **Probability:** $x(1-x) \approx 24\%$
Event: mother gave M_- **Probability:** x	**Genotype:** M_-M_+ **Trait:** dark eyes **Probability:** $x(1-x) \approx 24\%$	**Genotype:** M_-M_- **Trait:** blue eyes **Probability:** $x^2 \approx 17\%$

$\diamondsuit$

2.4.2 Using Population Statistics in Probability Computations

If populaiton statistics is available, we can revisit our old example (introduced on page 18) about a family with one blue eyed, one dark eyed parent, and one dark eyed child.

Assuming that this family is American, we can use the data from the above Punnett square to compute the $P\left(\begin{array}{c}\text{dark eyed}\\ \text{parent is}\\ \text{homo-}\\ \text{zygous}\end{array}\right) = p$ used in that old example:

$$P\left(\begin{array}{c}\text{dark eyed}\\ \text{parent is}\\ \text{homo-}\\ \text{zygous}\end{array}\right) = P\left(\begin{array}{c}\text{American}\\ \text{is}\\ M_+M_+\end{array}\middle|\begin{array}{c}\text{American}\\ \text{is dark}\\ \text{eyed}\end{array}\right) =$$

$$= \boxed{\text{expressing conditional formula in terms of unconditional}} =$$

$$= \frac{P\left(\begin{array}{ccc}\text{American} & & \text{American}\\ \text{is} & \bigcap & \text{is dark}\\ M_+M_+ & & \text{eyed}\end{array}\right)}{P\left(\begin{array}{c}\text{American}\\ \text{is dark}\\ \text{eyed}\end{array}\right)} = \boxed{\begin{array}{ccc}\text{American} & & \text{American}\\ \text{is} & \subseteq & \text{is dark}\\ M_+M_+ & & \text{eyed}\end{array}} =$$

$$= \frac{P\left(\begin{array}{c}\text{American}\\ \text{is}\\ M_+M_+\end{array}\right)}{P\left(\begin{array}{c}\text{American}\\ \text{is dark}\\ \text{eyed}\end{array}\right)} = \boxed{\text{use the Punnett square}} =$$

$$\approx \frac{35\%}{83\%} \approx 42\%.$$

Now we can use the formula obtained on page 18 with the actual value of p that is true for the American population:

$$P\left(\begin{array}{c|c}\text{dark eyed} & \text{child}\\ \text{parent is} & \text{is}\\ \text{homo-} & \text{dark}\\ \text{zygous} & \text{eyed}\end{array}\right)=\boxed{\text{old formula}}=$$

$$=\frac{2p}{p+1}=\boxed{\text{using the value of } p \text{ that was just computed}}=$$

$$=\frac{2\cdot 42\%}{42\%+1}\approx 59\%.$$

Thus, if an American family has one blue eyed and one dark eyed parent, as well as a dark eyed child, then the probability of the dark-eyed parent being homozygous is approximately 59%.

References

[1] Bayes, Thomas and Price, Richard. "An Essay towards Solving a Problem in the Doctrine of Chances". In: *Philosophical Transactions of the Royal Society of London* 53 (1763). By the Late Rev. Mr. Bayes, F. R. S. Communicated by Mr. Price, in a Letter to John Canton, A. M. F. R. S, pp. 370–418.

[2] Douglas Belkin. *Blue eyes are increasingly rare in America.* URL: https://www.nytimes.com/2006/10/18/world/americas/18iht-web.1018eyes.3199975.html. The New York Times, Published October 18, 2006; Retrieved May 11, 2020.

[3] *Double Helix.* URL: https://www.genome.gov/genetics-glossary/Double-Helix. National Human Genome Research Institute, Retrieved April 1, 2020.

[4] Gregor Johann Mendel. "Versuche über Pflanzenhybriden [Experiments in plant hybridization]". In: *Verhandlungen des naturforschenden Vereines in Brünn [Proceedings of the Brünn Natural History Society]* IV für das Jahr 1865 (1866), pp. 3–47.

[5] *Saving Women's Lives: Strategies for Improving Breast Cancer Detection and Diagnosis.* URL: https://www.ncbi.nlm.nih.gov/books/NBK22311/. National Center for Biotechnology Information, NIH, Retrieved April 21, 2020.

[6] *Serology-based tests for COVID-19.* URL: https://www.centerforhealthsecurity.org/resources/COVID-19/serology/Serology-based-tests-for-COVID-19.html. Center for Health Security, Johns Hopkins Univertity Bloomberg School of Public Health, Retrieved April 21, 2020.

Index of Terms

Index of People