Chapter 1

Probability and Combinatorics

Chance is what makes life worth living. If everything was known in advance imagine the disappointment.

Dr. J.K. Das

Random nature of observed reality

Why *the same* parents could have so *different* children? Why *the same* coin being tossed two times provides *different* results?

The answer is that, even under apparently similar conditions, there is a large number of uncontrollable (sometimes unknown) variables that we do not measure. These numerous variables have cumulative effect, creating variation in the observed result. It is this variation that we term *randomness*. Think about tossing a coin. It will come up heads or tails depending on a number of factors such as speed of tossing a coin, the height from which it falls, the wind, etc. If we could accurately measure and take into account all these factors, we would be able to predict the result of each tossing exactly. However, since we cannot instantly do all this complicated work we simply view the result as random!

Here we present an introduction to the theory which accounts for this "random" component via the concept of probability. For example, having no instruments to predict the result of a tossing, we say that a coin comes up heads with probability $\frac{1}{2}$.

Then, if you know the true nature of things, nothing is random. Though, as long as you are a human being, you need probability theory to make sense of and navigate yourself in this complicated world.

Probability

Definition of probability

Random Experiment is a process leading to two or more possible outcomes without knowing exactly which outcome will occur. Rolling a die is the an example of experiment.

The possible outcomes from a random experiment are called the **basic outcomes** **(O_i)**, and the set of all basic outcomes is called the **sample space** **(S)**. Note that basic outcomes are the smallest possible indivisible elements of the sample space which cannot be divided into "more elementary" outcomes. Example of a basic outcome for rolling a die is that it will face up "4". The sample space is the set: $\{1, 2, 3, 4, 5, 6\}$.

Event is any subset of interest of basic outcomes from the sample space, it can be denoted by a letter. For example "a die will show up more than 3" is as an event and can be denoted by A. The **null event** corresponds to the situation when none of the basic outcomes happens, it is an impossible event. For example "a die will show up 7" is a null event. It is denoted by $\emptyset$ (empty space).

In a very general sense probability is a quantifiable measure of the degree of belief in a particular event. So, if you don't expect to fail an exam, you would say that there is a low probability of this event. There are several approaches to calculating probability.

Classical probability of an event is the ratio of number of outcomes that constitute an event A (N_A) to all possible outcomes in the sample space (N), assuming that all outcomes are equally likely to occur.

$$P(A) = p = \frac{N_A}{N}$$

Again, in the example with rolling a die there are 6 equally likely basic outcomes, $N = 6$ (each with probability $\frac{1}{6}$). If you define event A as "a die will show up more than 3 dots", then $N_A = 3$ (there are 3 basic outcomes: "4", "5" and "6" which fit the event). Thus, $P(A) = \frac{3}{6} = 0.5$.

Note that probability of an event is the same as the *proportion* of times it occurs in experiments, it is conventionally denoted by letter p.

However, this type of calculation requires you to deeply know the "mechanism" of the experiment. Imagine that you want to calculate the probability that a randomly taken citizen will support "Edinaya Rossiya" party at the upcoming elections. First you'll need to discover all the elementary outcomes, or "bricks", which together build up political preferences of a person. Then, you have to identify which of them constitute the desire to vote for "Edinaya Rossiya". Social sciences do not provide us with such precise knowledge on the topic. So, we may never really calculate the classical probability for this type of events.

Therefore, in practice the probability of an event A (empirical probability) is estimated as its **relative frequency**. It is the ratio of times that event A occurred (n_A) in n number of trials.

$$P(A) \approx \hat{p} = \frac{n_A}{n}$$

For example, if you roll a die 100 times, and it shows up "4", "5" or "6" 53 times, then the estimated probability of the event "more than 3" is $\frac{53}{100} \approx 0.53$, which is not equal to the true value of 0.5.

The problem is that relative frequency is calculated based on the random result of some number of trials. If you repeat the same n number of trials you can reasonably get somewhat different number of times n_A event A occurred and thus the different estimated probability.

The letter $\hat{p}$ here denotes the estimated or sample proportion of times the event occurs in the experiments.

The **law of large numbers** states that when an experiment is conducted a large number of times relative frequency of an event will converge to its classical (true) probability.

To demonstrate the law of large numbers let's say I flip a coin 3 times and occasionally each time it comes up heads. For this experiment the empirical probability of the event "heads" is $\frac{3}{3} = 1$.

However, if I were to flip the coin 100 times I would expect the empirical probability of this experiment to be much closer to the true probability of 0.5.

That's why the true probability is sometimes named a **limiting relative frequency**.

Properties of probability

1. Probability of an event can be calculated as the sum of probabilities of all its basic outcomes.

$$P(A) = P(O_1) + P(O_2) + \ldots + P(O_{N_A})$$

2. Probability of any event is between 0 (event never occurs in an experiment) and 1 (event always occurs in an experiment). The larger is the probability, the more likely is the event.

$$0 \leq P(A) \leq 1$$

3. Sum of probabilities of all basic outcomes O_i (along the sample space S) is 1.

$$P(O_1) + P(O_2) + \ldots + P(O_N) = P(S) = 1$$

4. Probability of the null event is zero.

$$P(\emptyset) = 0$$

Operations on events

Diagrams below are called Venn diagrams, which are very useful tools in probability analysis. The rectangle represents the sample space S and the circles in it represent events. Probability of an event is the area of the corresponding circle, the area of the whole rectangle (sample space) is 1.

- **Intersection** of A and B is the event when both A *and* B occur. It is denoted by $A \cap B$. Probability of $A \cap B$ is called **joint probability** of A and B. It is shaded green on the diagram below.

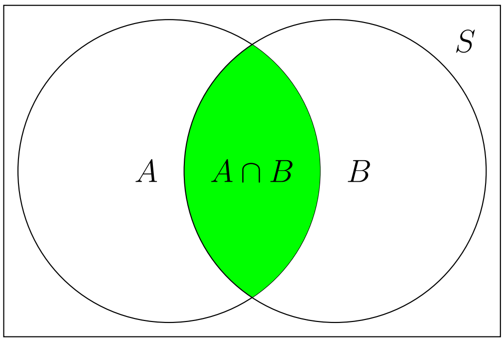


Figure 1.1: Intersection of A and B

Consider an experiment of randomly choosing an ICEF student. Let $A =$ "student's name starts with A" and $B =$ "student is a boy". Then, $A \cap B$ is an event that a chosen student is a boy *and* his name starts with A. For example, "Alexey" is one of the basic outcomes belonging to $A \cap B$.

- If events A and B have no intersection (can never occur simultaneously) they are called **mutually exclusive** or **disjoint**.

$$A \cap B = \emptyset, \; P(A \cap B) = 0$$

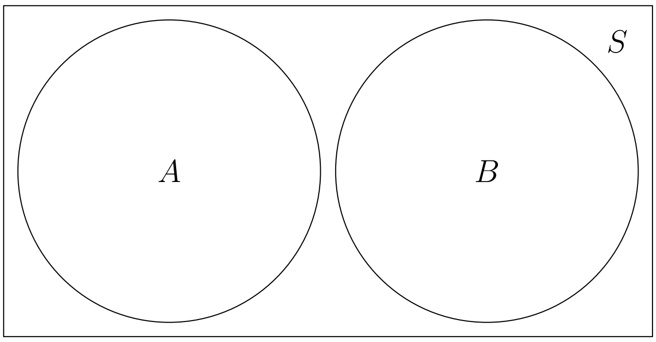


Figure 1.2: For example $A =$ "Anastasia" and $B =$ "a boy".

- **Union** of A and B, denoted by $A \cup B$, is the event when at least one of them occur. $A \cup B$ happens when A *or* B or both A and B happen.

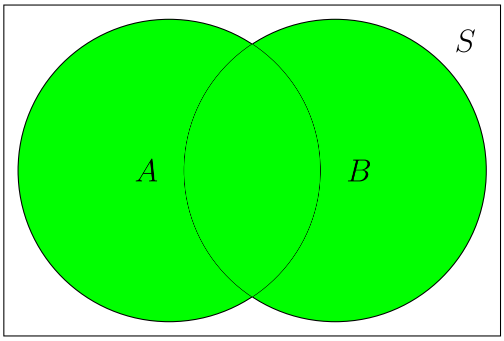


Figure 1.3: Union of events

For A = "name starts with A" and B = "a boy" $A \cup B$ corresponds to the case when either of the events occurred: a student chosen happened to be a boy *or* his/her name starts with A, *or* both events have happened: he is a boy with "A" name. Both "Anastasia" and "Alexey" belong to $A \cup B$, as well as their friend "Borya".

- **Complement** of A, denoted as $\bar{A}$ or A^c, is an event that A does not occur. $\bar{A}$ is read as "not A". It includes all the basic outcomes in S, which do not belong to A. The probability that an event will *not* happen is equal to 1 minus the probability that it will happen.

$$P(\bar{A}) = 1 - P(A)$$

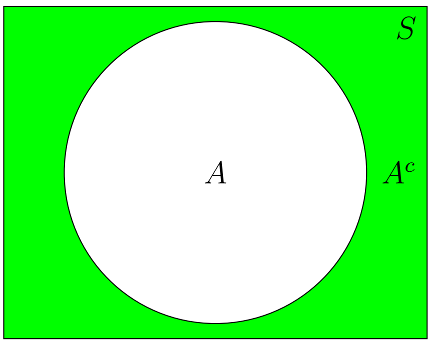


Figure 1.4: $\bar{A}$ is complement to A

For A = {"name starts with A"}, $\bar{A}$ is read as "not A" and happens whenever the name of a student starts with any other letter. Here "Borya" belongs to $\bar{A}$.

- Events are called **collectively exhaustive** if they are mutually exclusive and when taken together they represent the entire sample space. In other words, such events *exhaust* the complete set of all possible outcomes.

$$E_i \cap E_j = \emptyset \text{ for all } i,\ j \text{ and } E_1 \cup E_2 \cup \ldots \cup E_k = S$$

$$P(E_1 \cup E_2 \cup \ldots \cup E_k) = P(E_1) + P(E_2) + \ldots + P(E_k) = P(S) = 1$$

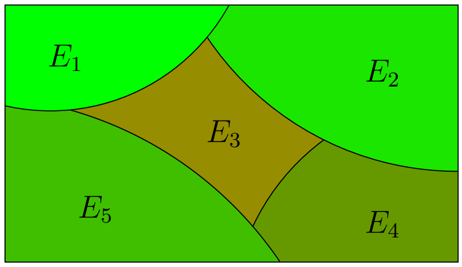
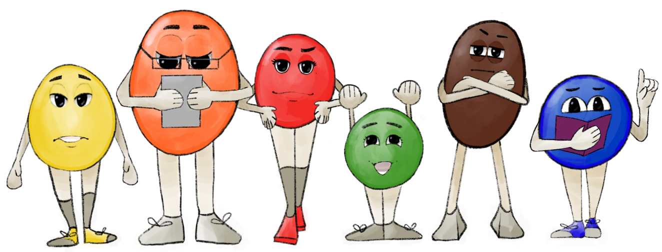


Figure 1.5: Collectively exhaustive events

For example, the set $E_A, E_B, \ldots, E_Z$ represents the set of events that a randomly chosen student has name starting with each possible letter from A to Z, correspondingly. Then, $E_A, E_B, \ldots, E_Z$ are collectively exhaustive.

The other example is the set of colors of M&M's candies. Events of randomly choosing a Yellow, Orange, Green, Red, Blue and Brown candy from a packing are collectively exhaustive.



Addition rule: probability of union

There is a formula for calculating probability of union of events:

$$P(A \cap B) = P(A) + P(B) - P(A \cap B).$$

Proof

As you can see on the diagram:

$$P(A) = x + y,\ P(B) = z + y,\ P(A \cup B) = y,\ P(A \cap B) = x + y + z.$$

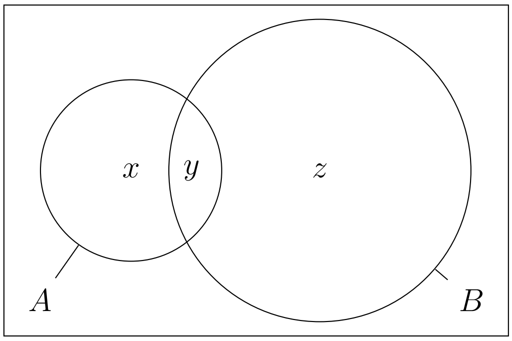


Figure 1.6: $P(A) = x + y,\ P(B) = y + z$

If you sum up $P(A)$ and $P(B)$ you get:

$$P(A) + P(B) = x + y + z + y = x + 2y + z \neq (A \cap B).$$

In order to get $x + y + z$ we should subtract y from $P(A) + P(B)$. Thus,

$$P(A \cup B) = P(A) + P(B) - y = P(A) + P(B) - P(A \cap B).$$

Conditional Probability

$A|B$ is an event that A will occur *given that B has already occurred.* Conditional probability $P(A|B)$ is read as probability of A given B.

Think of rolling a die. Let event A be "a die faces up 6", $B =$ "a die faces up with an even number". It is obvious that $P(A) = \frac{1}{6}$ and $P(B) = \frac{3}{6} = \frac{1}{2}$.

Then, $P(A|B)$ denotes the probability that the die faced up "6", in the case that you already know that the number occurred is even. Thus, we need to estimate probability of event "6" from the set of all even numbers $\{2, 4, 6\}$, and not from the sample space $\{1, 2, 3, 4, 5, 6\}$ as initially. Thus, $P(A) = \frac{1}{6}$ and $P(A|B) = \frac{1}{3}$. As you can see, $P(A|B)$ is not the same as $P(A)$.

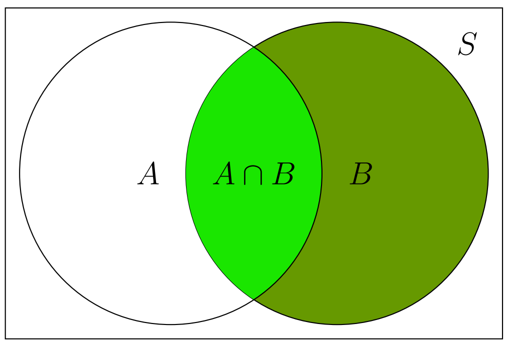


Figure 1.7: Conditional Probability

Conditional probability addresses the idea that *once event B has happened, the chance of A changes!* Initially (before B occurred) the sample space was represented by S as denoted on the diagram. Then B has occurred. Probability of a happened event is 1, so, the sample space shrinked to the area of B and thus A can only occur as a part of B. From this point A is only represented by intersection of A and B and should be calculated as a part of the new probability space, that is — as a part of B.

Thus, we arrive to the definition of **conditional probability**:

!

$$P(A|B) = \frac{P(A \cap B)}{P(B)}.$$

Why the fact that one event has occurred changes the probability of the other event? Think of the following. If event A is to "meet a shark", in your everyday life the probability of event A is close to zero. Indeed, how is it possible to meet a shark somewhere in the center of Moscow? Now denote event B as "being in the deep sea" and assume that B has happened. The situation has changed completely! Given B, you don't feel that safe anymore: the possibility to meet a shark does not seem a nonsense now. Here $P(A) \approx 0$, but $P(A|B)$ is significantly larger than 0.

Some hints:

1. From definition of conditional probability we also get:

$$P(A \cap B) = P(A|B) \cdot P(B) = P(B|A) \cdot P(A).$$

2. Note that for disjoint events:$A \cap B = \emptyset$, $P(A \cap B) = 0$ and additive law simplifies to:

$$P(A \cup B) = P(A) + P(B).$$

3. We can also express $P(A \cap B)$ from the additive law and use the formula:

$$P(A \cap B) = P(A) + P(B) - P(A \cup B).$$

4. Note that $A \cup B$ can be read as "at least one of the events A and B happens". Its probability corresponds to the green space on figure 1.3. All that remains — white area on the diagram, — represents cases when *none* of A and B happens, that is $P(\bar{A} \cap \bar{B})$. Sometimes it is easier to calculate $P(A \cup B)$ using probability of its complement:

$$P(\text{“at least one happens”}) = 1 - P(\text{“none happens”})$$

or

$$P(A \cap B) = 1 - P(\overline{A \cup B}) = 1 - P(\bar{A} \cap \bar{B})$$

Total (Full) Probability formula

A cheating student. Consider an example. First year ICEF students are writing an exam on Statistics. Two rooms are chosen: a big one and a small one. The first room is big — it contains 60% of all students writing the exam, and 5% of them are trying to cheat. The second room is small, it contains 40% of all students, and 2% of them are trying to cheat.

If a student is chosen at random what is the probability that he is a *cheater*?

The Full Score Strategy: problems on probability.

While solving this problem we give an illustration of the standart steps that need to be present in your solution to be marked with a full score.

Step 1. Sing the "Let it be song" — introduce your variables ("Let A be the event of...")

Let A be "a cheating student", B — "student sitting in the big (first) examination room" Then, $\bar{B}$ is "student was in the small room".

Step 2. Write down all the given probabilities in terms of variables introduced.

We have: $P(B) = 0.6$, $P(A|B) = 0.05$, $P(\bar{B}) = 0.4$, $P(A|\bar{B}) = 0.02$

Step 3. Write what probability you want to find. $P(A)$-?

Step 4. Write down the formula.

Obviously, a cheater can be found in each room. In other words, A consists of its intersection with B and its intersection with $\bar{B}$. Thus, $P(A) = P(A \cap B) + P(A \cap \bar{B})$. Using **hint 1** we get the *formula of total probability:*

!

$$P(A) = P(A|B) \cdot P(B) + P(A|\bar{B}) \cdot P(\bar{B}).$$

Step 5. Do the calculations and produce the answer to the question.

Thus, probability to randomly find a cheater is: $P(A) = 0.05 \cdot 0.6 + 0.02 \cdot 0.4 = 0.038$. Thus, a randomly chosen student cheats with probability of 0.038.

In this formulation of the total probability we look at the event A in the context of two conditions — sitting in the big or the small rooms, B and $\bar{B}$. However, there can be more than 2 circumstances ("states of nature") for A. Let $B_1, B_2, \ldots, B_n$ be a set of collectively exhaustive events. Using the same logic as above we can represent probability of A as:

$$P(A) = P(A|B_1) \cdot P(B_1) + P(A|B_2) \cdot P(B_2) + \ldots + P(A|B_n) \cdot P(B_n)$$

.

Bayes' rule

From the definition of conditional probability:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \text{ and } P(B|A) = \frac{P(A \cap B)}{P(A)}$$

Thus,

$$P(A \cap B) = P(A|B) \cdot P(B) = P(B|A) \cdot P(A)$$

Hence,

$$P(B|A) = \frac{P(A|B) \cdot P(B)}{P(A)}$$

This is the first formula for the Bayes' rule. Applying the full probability formula to $P(A)$ we come to the second formulation of the Bayes rule:

!

$$P(B|A) = \frac{P(A|B) \cdot P(B)}{P(A|B) \cdot P(B) + P(A|\bar{B}) \cdot P(\bar{B})}.$$

Example: Inverse to "a cheating student" problem. A randomly selected student turned out to be a cheater. What is the probability that he was seating in the first room?

Using Bayes' rule:

$$P(B|A) = \frac{P(A|B) \cdot P(B)}{(A|B) \cdot P(B) + P(A|\bar{B}) \cdot P(\bar{B})} = \frac{0.05 \cdot 0.6}{0.05 \cdot 0.6 + 0.02 \cdot 0.4} = 0.789$$

Independent events

Two events A and B are **statistically independent** if and only if:

$$P(A \cap B) = P(A) \cdot P(B)$$

From the first hint we know that $P(A \cap B) = P(A|B) \cdot P(B) = P(B|A) \cdot P(A)$. From that, we also get that $P(A|B) = P(A)$ and $P(B|A) = P(B)$. Thus, for independent events conditional probability equals unconditional probability. This simply means that when A and B are independent the fact that B happened does not influence the chance of A, and vice versa.

For example, the fact that some "babushka" Maria Petrovna is having her 73$^{\text{rd}}$ birthday today influences your chance of staying in a good mood today (if Maria Petrovna is not your grandmother).

In the case of more than two events we call the set of $A_1, A_2, \ldots, A_n$ as mutually independent if for any subset of them $A_{K_1}, A_{K_2}, \ldots A_{Km}$ it holds that: $P(A_{K_1} \cap \ldots \cap A_{K_1}) = P(A_{K_1}) \cdot \ldots \cdot P(A_{Km})$.

Probability tree

There is a very useful tool for solving problems on probability with more than one event, called probability tree. See how "a cheating student" problem can solved be solved using this tool:

The upper "layer" of the diagram represents probabilities of B and $\bar{B}$. The lower one coming from B and $\bar{B}$ represents the corresponding *conditional* probabilities of A and $\bar{A}$. This tree provides a nice illustration for the following calculations:

$P(A \cap B) = 0.6 \cdot 0.05 = 0.03$: look where both A and B happen on the diagram — the left branch. Then, multiply probabilities on it.

$P(A) = 0.05 \cdot 0.6 + 0.02 \cdot 0.4 = 0.038$: look where A happens on the diagram: letters A occur on the two branches of the tree. Find the probability of A on each branch and then sum it up.

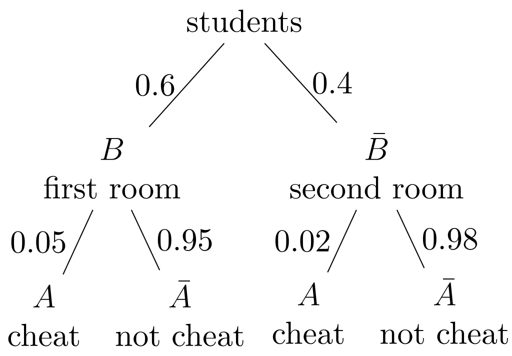
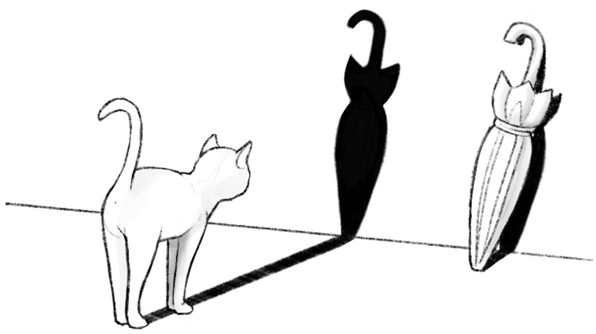


Figure 1.8: solving the "cheating student" problem with a probability tree

In many problems probability tree allows to get an answer in a simple and intuitive way.

Note that probability of intersection ($A \cap B =$ "student is a cheater and from the first room") and conditional probability ($A|B =$ "student chosen from the first room is a cheater") are not the same! Although the resulting outcomes seem to be equal, they address to different situations. $P(A \cap B) = 0.03$ is a chance that a student chosen at random from all first year students is both a cheater and writes the exam in the first room. $P(A|B) = 0.05$ is a chance that, being in the first examination room you will randomly choose a cheater. Once again: remember that probabilities written on the branches of probability tree are conditional. This is very intuitive feature since the branch is coming from the event (B), which addresses the idea of analyzing probabilities of events inside the new sample space, were this event has already happened (given B). For more information about this distinction see the "Top secret information" section at the end of this chapter.



Combinatorics

Let us remind you, that according to the classical definition probability is the ratio of number of outcomes that pertain a random event A (N_A) to the number of all possible outcomes in the sample space (N), given that all outcomes are equally likely to occur.

$$P(A) = \frac{N_A}{N}$$

Thus, to calculate the probability we need to count the number of possible outcomes N and N_A. Combinatorics is a branch of mathematics which provides a tool for such calculations.

Number of ordered sequences with repetition
Think of a game with n possible outcomes in each round (if the game is rolling a die, then $n = 6$). There are k independent rounds in the game (a die is rolled k times). How many different results can the game have? The first round can bring n different outcomes, the second round can also bring n outcomes. So, the two rounds can lead to $n \times n$ different possible results of the whole game. Since in each round there are n possible outcomes, each round raises the set of possible results by n. So, in k rounds we come to n^k possible results.

The number of possible results (sequences of outcomes after the k rounds) is in fact the number of possible ordered sequences of size k constructed of n objects.

$$n^k$$

Example. If you roll a die 5 times, there can be $6^5 = 7776$ different results (sets of 5-digit sequences of numbers from 1 to 6).

Number of ordered sequences of k objects without repetition. Permutations of k objects.
Suppose you have a set of k students waiting to enter the classroom to pass the oral examination on Philosophy. Students come to the room one after another. How many different orderings of students (or simply put, lists of students' names) can be arranged of k? The first student to come can be any of k students. Once he/she entered the room there are $(k-1)$ students left. Thus, the second student can be any of $(k-1)$ students. The third — any of $(k-2)$, etc. up to the moment when the last student is left, and he can be the only one to finally enter the room. Thus, overall number of permutations is $k \cdot (k-1) \cdot (k-2) \cdot \ldots \cdot 2 \cdot 1$ which is a factorial of k, denoted as $k!$. Note, that $0! = 1$.

$$k! = k \cdot (k-1) \cdot (k-2) \cdot \ldots \cdot 2 \cdot 1$$

Example. In a class of $k = 10$ students there can be $10! = 3628800$ possible orderings.

Number of ordered sequences of k out of n without repetition. Permutations of k objects out of n. Now suppose there are n students in the class waiting to pass the exam. But only k of them will be examined today, $k < n$. How many different orderings can be created out of k students taken from the class of n students? Again, the first student to enter can be anyone of n students, the second — any of $(n-1)$, etc. However, when k students are examined the exam is stopped, with $(n-k)$ students staying unexamined. Thus, the last student to attempt the exam is anyone out of $(n-k+1)$ students. Thus, overall number of permutations is $n \cdot (n-1) \cdot \ldots \cdot (n-k+1)$. If we multiply and divide it by $(n-k)!$ we will get $\dfrac{n \cdot (n-1) \cdot \ldots \cdot (n-k+1) \cdot (n-k)!}{(n-k)!} = \dfrac{n!}{(n-k)!}$

Permutations of k out of n are generally denoted by A_n^k (Russian style) or P_n^k (English style).

$$A_n^k = P_n^k = \frac{n!}{(n-k)!}$$

Example If $k = 6$ students are to be examined from a class of $n = 10$ the number of possible ordering is: $P_{10}^6 = \frac{10!}{(10-6)!} = \frac{10!}{4!} = \frac{3628800}{24} = 151200$.

Number of unordered sequences of k out of n without repetition. Combinations of k out of n.

Now suppose that you are not interested in different orderings of the same students. You only want to know *who* will be able to attempt the exam today. So, you want to find the number of different sets of k students taken from the class of n. You've already learned that there can be $\frac{n!}{(n-k)!}$ different permutations of k students. Each of the sets of k chosen students can be ordered in $k!$ different ways. So, the number $\frac{n!}{(n-k)!}$ takes into account all permutations of each possible set of k students. To get the number of unique sets it should be lowered by the factor $k!$. Thus, the number of different sets (combinations) of k students taken from the class of n equals to $\frac{n!}{k! \cdot (n-k)!}$. Combinations of k out of n are generally denoted as C_n^k or $\binom{n}{k}$.

$$C_n^k = \text{problem} = \frac{n!}{k! \cdot (n-k)!}$$

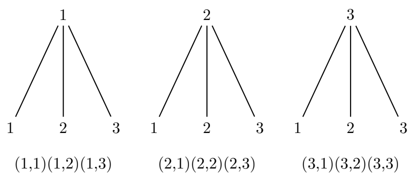
Example. If $k = 6$ students are to be examined from a class of $n = 10$ the number of possible sets (groups) of students to be examined today is: $C_{10}^6 = \dfrac{10!}{6! \cdot (10-6)!} = \dfrac{10!}{6! \cdot 4!} = \dfrac{3628800}{720 \cdot 24} = 210$

Note that $C_n^k = C_n^{n-k}$. Prove it yourself, using the formula above. The intuition here is obvious: choosing k students to attempt the exam today is the essentially the same as choosing $(n-k)$ students (the rest) who will not attempt the exam.

Get to know the first year student Masha who was introduced to you in the Intro part of the book. She is startled by combinatorics, let's help her to make sense of this theory! Consider a set $\{1, 2, 3\}$. It represents the three possible outcomes from one round of a "game", $n = 3$. What do they mean by ordered and unordered sequences with or without repetition?

1. Ordered[2] sequences with repetition[3]. n^k

   Suppose a number is chosen from the set 2 times. This is like 2 rounds of a game with 3 outcomes: $n = 3$, $k = 2$. How many different results (sequences of numbers) can this game provide? Using the formula we get: $n^k = 3^2 = 9$. What are the 9 possible results of this game? You can simply list them to make sure the answer is true. This is the tree of possible outcomes:

   As you can see, overall it shows 9 sequences:

(1,1)	(2,1)	(3,1)
(1,2)	(2,2)	(3,2)
(1,3)	(2,3)	(3,3)

2. Permutations of k objects (ordered sequences without repetition) $k!$ In how many ways can we order the numbers in the set {1,2,3}? The formula suggests there are $k! = 3! = 6$ ways to do that. Here they are:

(1,2,3)	(2,1,3)	(3,1,2)
(1,3,2)	(2,3,1)	(3,2,1)

[2] Ordered sequence means that order matters in this counting: (1,2) and (2,1) are viewed as different sequences

[3] With repetition means that the same outcomes may repeat in different rounds: results (1,1) or (3,3) are possible.

3. Permutations of k objects out of n (ordered sequences without repetition) P_n^k In how many ways can we order 2 numbers chosen from the set {1,2,3}? $n = 3$, $k = 2$. By the formula there are 6 possible outcomes in this game:

$$P_3^2 = \frac{3!}{(3-2)!} = \frac{3!}{1!} = 3! = 6$$

It is easy to define these 6 sequences:

(1,2) (2,1) (3,1)
(1,3) (2,3) (3,2)

4. Combinations of k out of n (unordered sequences of k out of n without repetition) C_n^k In how many ways can we choose two numbers from the set $\{1, 2, 3\}$, if the same numbers in different order are viewed as the same sequence? $n = 3$, $k = 2$. By the formula there are three possible outcomes in this game:

$$C_3^2 = \frac{3!}{(3-2)! \cdot 2!} = \frac{3!}{1! \cdot 2!} = 3$$

Here they are: (1,2), (1,3), (2,3).

Full score strategy

Note that the 5 steps to solve a problem provided on page 8 give a universal strategy for solving problems on probability. Try to follow it in your solutions to get the full mark:
Step 1. "Let it be" song — introduce your variables.
Step 2. Write down given probabilities.
Step 3. Write the probability you want to find.
Step 4. Provide the formula.
Step 5. Make calculations and the give answer.

You MUST BE ABLE TO REPRODUCE even being half-awake:

- The law of large numbers: when an experiment is conducted a large number of times the observed relative frequencies of an event will converge to its classical (true) probability.
- $0 \leq P(A) \leq 1$
- Statistical independence:

$$P(A \cap B) = P(A) \cdot P(B) or P(A|B) = P(A)$$

- Full probability formula:

$$P(A) = P(A|B_1) \cdot P(B_1) + P(A|B_2) \cdot P(B_2) + \ldots + P(A|B_n) \cdot P(B_n)$$

- Bayes rule: $P(B|A) = \dfrac{P(A|B) \cdot P(B)}{P(A)}$

- Permutations: $P_n^k = \dfrac{n!}{(n-k)!}$

- Combinations: $C_n^k = \dfrac{n!}{k! \cdot (n-k)!}$

- Operations on events. Summary.

Term	Description	Symbol	Properties
Intersection of	AND/BOTH	$A \cap B$	$P(A \cap B) = P(A) + P(B) - P(A \cup B)$
Union of	OR/ at least one	$A \cup B$	$P(A \cup B) = P(A) + P(B) - P(A \cap B)$
Complement	NOT	$\bar{A}$ or A^C	$P(\bar{A}) = P(A^C) = 1 - P(A)$
Conditional on	GIVEN	$A\|B$	$P(A\|B) = \dfrac{P(A \cap B)}{P(B)}$
Mutually exclusive	AT MOST one		$A \cap B = \emptyset,\ (A \cap B) = 0$
Collectively exhaustive	Taken together exhaust all possibilities		$E_i \cap E_j = \emptyset$ and $E_1 + E_2 + \ldots + E_k = S$

Calculator BOX

For simple combinatorics calculations use the one of the following routes in your calculator:

- `RUN-MAT` $\rightarrow$ `OPTN` $\rightarrow$ `PROB`
- `STAT` $\rightarrow$ `OPTN` $\rightarrow$ `PROB`

Then use:

- `x!` to calculate factorial
- `nPr` to calculate the number of permutations of r objects out of n
- `nCr` to calculate the number of combinations of r objects out of n

Top secret information

Here is the problem, guys.

Probability that a coin tossed two times will come up heads on the first and on the second trial is $\frac{1}{4}$ (calculated as $\frac{1}{2} \cdot \frac{1}{2}$).

Probability that a coin which came up heads on the first trial, will come up heads on the second trial again is $\frac{1}{2}$ (trials are independent, so probability is the same as the probability of heads in a single toss).

Both situations result in the event "two heads". Why having different probabilities?!

Here are the few recipes to overcome this *existential crisis.*[4]

Let H_1 and H_2 denote the events when a coin comes up heads on the first and on the second trial, correspondingly. S is the sample space. We know that: $P(H_1) = P(H_2) = \frac{1}{2}$, $P(S) = 1$. We need to explain why $P(H_2|H_1)$ is not the same as $P(H_1 \cap H_2)$. In some explanations we will use that: $P(H_1 \cap H_2) =$ [since H_1 and H_2 are independent] $= \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$

By the classical definition of probability

By definition the true probability of an event is the proportion of the number of outcomes that pertain the event to the number of all possible outcomes, $\frac{N_A}{N}$.

Let's also introduce the event T, which is coin coming up tails. The event $H_1 \cap H_2 = HH$ is taken from the "game" with tossing a coin two times. There are 4 possible outcomes, which are: HH, HT, TH, and TT. Thus, $N = 4$. Only one of these equiprobable outcomes corresponds to the event, $N_A = 1$. Thus, $P(H_1 \cap H_2) = \frac{1}{4}$.

Contrary, event $H_2|H_1$ is from the "game" with tossing a coin only once. Of course, we have knowledge that on the previous trial the coin came up heads, but we proceed with tossing a coin only once. So, there are 2 equally likely outcomes: H and T, $N = 2$. Only H satisfies the event of interest, $N_A = 1$. Thus, $P(H_1|H_2)$

Graphically

Below is the sample space S for this "game". Event $H_1 \cap H_2$ is represented by the intersection of the two circles inside the rectangle of the sample space. Therefore, its probability is equal to $\frac{1}{4}$ over 1. Here 1 is the probability of the sample space, from which the event $H_1 \cap H_2$ is "drawn". Thus,

$$P(H_1 \cap H_2) = \frac{\frac{1}{4}}{1} = \frac{1}{4}$$

[4]An *existential crises* is a moment at which an individual questions the very foundations of their life: whether this life has any meaning, purpose, or value (Richard K. James).

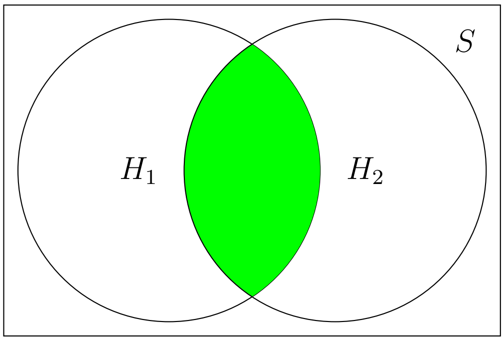


Figure 1.9: Intersection of H_1 and H_2

Event $H_2|H_1$ means that H_1 has already happened, and the sample space is now contracted to the area of H_1 circle.

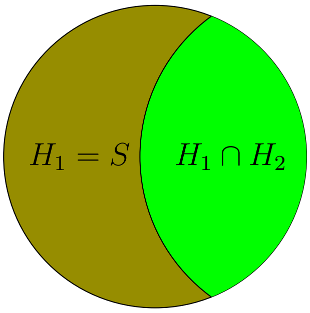


Figure 1.10: Conditional probability

Therefore $H_2|H_1$ is represented by the intersection (shaded area) as a part of H_1, which is the new sample space. Its probability is equal to $\frac{1}{4}$ over $\frac{1}{2}$. $P(H_2|H_1)=\frac{1/4}{1/2}=\frac{1}{2}$.

By the formula

By the definition of conditional probability we have: $P(H_2|H_1) = P\frac{(H_1 \cap H_2)}{P(H_1)}$. Again, we get $P(H_2|H_1) = \frac{1/4}{1/2} = \frac{1}{2} \neq P(H_1 \cap H_2) = \frac{1}{4}$.

Using properties of independent events

Since H_1 and H_2 are independent, $P(H_2|H_1) = P(H_2) = \frac{1}{2} \neq P(H_1 \cap H_2) = \frac{1}{4}$ Why is this the case? Remember, coins have no memory. Each time you toss a coin it just fails heads or tails equally likely, never mind what you've observed previously.

The infinity trick

One of my friends once asked me: if you toss a coin and it comes up heads four times in a row, would you bet on tails or heads on the next tossing. "It does not matter", — I answered, "since the chances of the two outcomes are always equal, no matter how many trials have been made". Then, he claimed the Law of Large Numbers is false.

Here are his arguments provided. By LLN I know that the relative frequency of tails in 10000 trials is closer to $\frac{1}{2}$, than in 1000 trials, which is still closer to $\frac{1}{2}$ than the relative probability in 100 trials. In other words, as number of trials increases, relative frequency approaches to one half, which means that the number of heads and tails should balance out in the long run and approach equality. Then, after four tosses resulted in heads the fifth toss should have larger probability of tails, so as to guarantee that the balance is approached in the long run.

This is the wrong logic. When we tell that in 10000 trials 5000 should result in heads, we talk about the expected, but not guaranteed result. In the matter of fact, it is possible that the coin will come up heads in all of 10000 trials (probability of that is small, but this is still possible). The law of large numbers works strictly only in the situation of infinite number of trials. Note, that infinity has never promised something to happen in this finite world of observed events. There is no "justice" for random events in the context of finite number of trials, a coin will not strive to fall up tails, thinking "I only have $(n-4)$ to restore the balance". A coin "does not care" about the balance in a finite number of trials n since she has an infinity in spare to restore that balance...

So, be careful, the law of large numbers works, but only as a limiting tendency.

You see how the pure randomness spilled into your life.

It is a question: do you toss a coin or does a coin toss you.

Sample AP problems with solutions

Problem 1. AP 2017 №3

You'll have about 13 minutes to solve this problem. It will bring you 15% of score for Free Response section.

A grocery store purchases melons from two distributors, J and K. Distributor J provides melons from organic farms. The distribution of the diameters of the melons from Distributor J is approximately normal with mean 133 millimeters (mm) and standard deviation 5 mm.

(a) For a melon selected at random from Distributor J, what is the probability that the melon will have a diameter greater than 137 mm? Distributor K provides melons from nonorganic farms. The probability is 0.8413 that a melon selected at random from Distributor K will have a diameter greater than 137 mm. For all the melons at the grocery store, 70 percent of the melons are provided by Distributor J and 30 percent are provided by Distributor K.

(b) For a melon selected at random from the grocery store, what is the probability that the melon will have a diameter greater than 137 mm?

(c) Given that a melon selected at random from the grocery store has a diameter greater than 137 mm, what is the probability that the melon will be from Distributor J?

Solution

(a) This probability is 0.2119. See Chapter 4.

(b) Let J and K denote that a melon is produced by Distributor J and K, correspondingly. Let G denote that a melon is greater than 137 mm in diameter. Thus, we know that: $P(G|K) = 0.8413$, $P(G|J) = 0.2119$. By the full probability formula:

$$P(G) = P(G|J)P(J) + P(G|K)P(K) = 0.2119 \cdot 0.7 + 0.8413 \cdot 0.3 \approx \mathbf{0.4007}$$

(c) By Bayes rule:

$$P\left(J|G\right) = \frac{P\left(G|J\right) \cdot P(J)}{P(G)} = \frac{0.2119 \cdot 0.7}{0.4007} \approx \mathbf{0.3701}$$

Note that the same answers for parts (b) and (c) can be derived using probability tree: *TBA*

Problem 2. AP 2014 №3

You'll have about 13 minutes to solve this problem. It will bring you 15% of score for Free Response section.

Schools in a certain state receive funding based on the number of students who attend the school. To determine the number of students who attend a school, one school day is selected at random and the number of students in attendance that day is counted and used for funding purposes. If more than 140 students are absent on the day the attendance count is taken for funding purposes, the school will lose some of its state funding in the subsequent year.

[...]

The principals' association in the state suggests that instead of choosing one day at random, the state should choose 3 days at random. With the suggested plan, High School A would lose some of its state funding in the subsequent year if the mean number of students absent for the 3 days is greater than 140.

[...]

(c) A typical school week consists of the days Monday, Tuesday, Wednesday, Thursday, and Friday. The principal at High School A believes that the number of absences tends to be greater on Mondays and Fridays, and there is concern that the school will lose state funding if the attendance count occurs on a Monday or Friday. If one school day is chosen at random from each of 3 typical school weeks, what is the probability that none of the 3 days chosen is a Tuesday, Wednesday, or Thursday?

Solution

Let A_1, A_2 and A_3 be events that the day chosen is Friday or Monday on the 1^{st}, 2^{nd} and 3^{rd} typical school weeks correspondingly.

$$P(\text{none of the 3 days chosen is a Tuesday, Wednesday, or Thursday}) =$$
$$= P(\text{all the 3 days chosen are either Friday or Monday}) = P(A_1 \cap A_2 \cap A_3)$$

Probability A_i that on a typical school week a randomly chosen day is Monday or Friday [e] is $\frac{2}{5} = 0.4$.

Since the days are chosen on the three weeks independently,

$$P(A_1 \cap A_2 \cap A_3) = P(A_1) \cdot P(A_2) \cdot P(A_3) = (\frac{2}{5})^3 = \mathbf{0.064}$$

[e] Note that here we calculate probability according to the classical definition: as number of possible outcomes representing the event divided by the number of all possible outcomes.

Problem 3. AP 2014 №2

You'll have about 13 minutes to solve this problem. It will bring you 15% of score for Free Response section.

Nine sales representatives, 6 men and 3 women, at a small company wanted to attend a national convention. There were only enough travel funds to send 3 people. The manager selected 3 people to attend and stated that the people were selected at random. The 3 people selected were women. There were concerns that no men were selected to attend the convention.

(a) Calculate the probability that randomly selecting 3 people from a group of 6 men and 3 women will result in selecting 3 women.

(b) Based on your answer to part (a), is there reason to doubt the manager's claim that the 3 people were selected at random? Explain.

Solution

(a)Event "randomly select 3 women" represents the intersection of three subsequent events: the first person selected is a woman, the second one is a woman and the third one is a woman. When selecting first, we choose out of 9 people, 3 out of whom are women. Thus, probability of the first event is $\frac{3}{9}$. Given that, next choice is made out of 8 people, 2 of whom are women, so, probability of the second event is $\frac{2}{8}$. After that, the third event has a $\frac{1}{7}$ chance to happen randomly. Thus, P("randomly select 3 women")$= \frac{3}{9}\frac{2}{8}\frac{1}{7} \approx \mathbf{0.012}$.

(b)In the beginning of the chapter we've told that in a very general sense probability is a quantifiable measure of the degree of belief in a particular event. Probability of only 1.2% indicates low degree of belief that it would happen under conditions of random selection. Then, if the selection was random, it would merely lead to choosing 3 women. Therefore, there is statistical *evidence to doubt* the manager's claim, that the selection was random.

Problem 4. AP 2011 №6

You'll have about 25 minutes to solve this problem. It will bring you 25% of score for Free Response section.

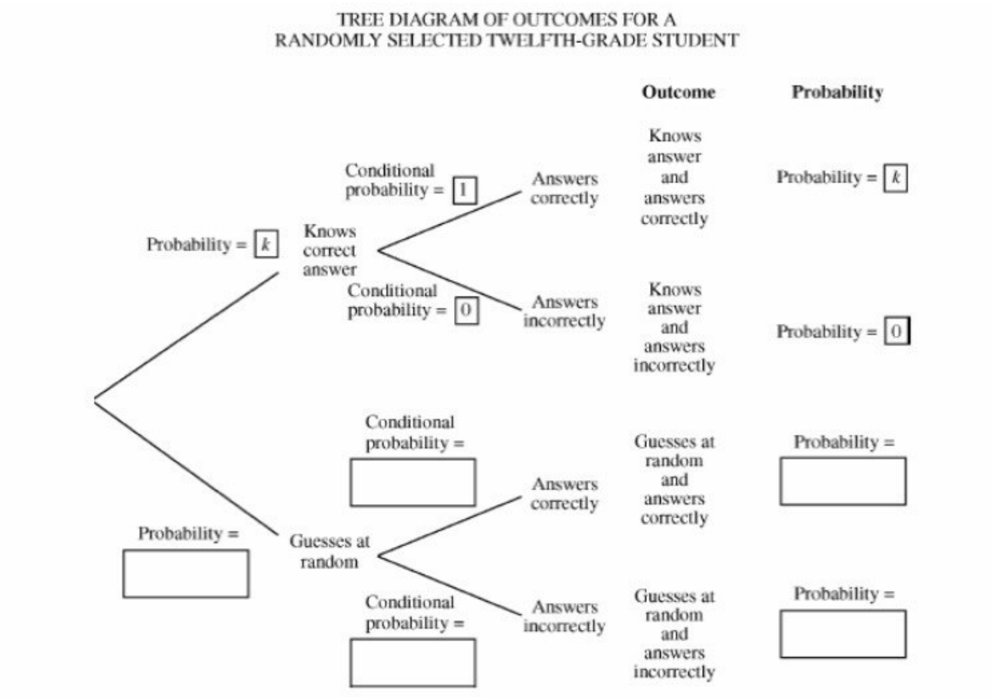
Every year, each student in a nationally representative sample is given tests in various subjects. Recently, a random sample of 9,600 twelfth-grade students from the United States were administered a multiple choice United States history exam. One of the multiple-choice questions is below. (The correct answer is C.)

> In 1935 and 1936 the Supreme Court declared that important parts of the New Deal were unconstitutional. President Roosevelt responded by threatening to
>
> (A) impeach several Supreme Court justices
>
> (B) eliminate the Supreme Court
>
> (C) appoint additional Supreme Court justices who shared his views
>
> (D) override the Supreme Court's decisions by gaining three-fourths majorities in both houses of Congress

Of the 9,600 students, 28 percent answered the multiple-choice question correctly.

Let p be the proportion of all United States twelfth-grade students who would answer the question correctly. Let k represent the proportion of all United States twelfth-grade students who actually know the correct answer to the question.

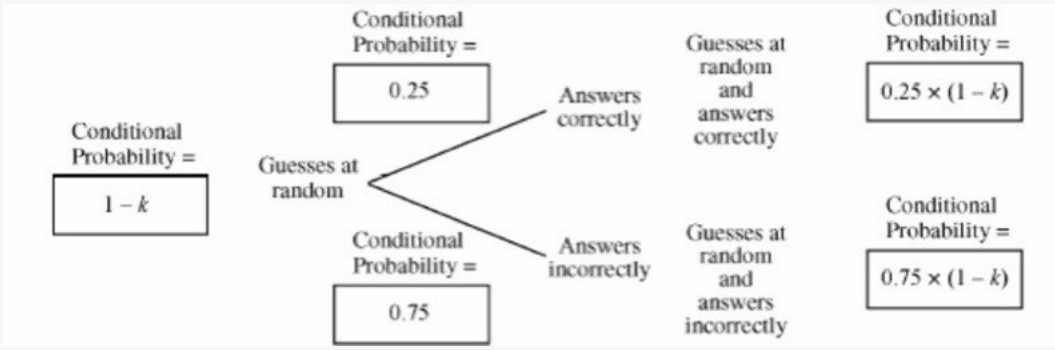
(b) A tree diagram of the possible outcomes for a randomly selected twelfth-grade student is provided below. Write the correct probability in each of the five empty boxes. Some of the probabilities may be expressions in terms of k.



(c) Based on the completed tree diagram, express the probability, in terms of k, that a randomly selected twelfth-grade student would correctly answer the history question.

Solution

(b) The five probabilities to be filled in the boxes are shown below.



(c) P(answers correctly) $= P$(knows correct answer and answerscorrectly)+ $+P$(guesses at random and answers correctly) $= k + 0.25(1 - k)$, which simplifies to $\mathbf{0.25 + 0.75}\boldsymbol{k}$.

Problem 5. AP 2009 Form B №2

You'll have about 13 minutes to solve this problem. It will bring you 15% of score for Free Response section.

The ELISA tests whether a patient has contracted HIV. The ELISA is said to be positive if it indicates that HIV is present in a blood sample, and the ELISA is said to be negative if it does not indicate that HIV is present in a blood sample. Instead of directly measuring the precence of HIV, the ELISA measures levels of antibodies in the blood that should be elevated if HIV is present. Because of variability in antibody levels among human patients, the ELISA does not always indicate the correct result.

As part of a training program, staff at a testing lab applied the ELISA to 500 blood samples known to contain HIV. The ELISA was positive for 489 of those blood samples and negative for the other 11 samples. As part of the same training program, the staff also applied the ELISA to 500 other blood samples known to not contain HIV. The ELISA was positive for 37 of those blood samples and negative for the other 463 samples.

(a) When a new blood sample arrives at the lab, it will be tested to determine whether HIV is present. Using the data from the training program, estimate the probability that the ELISA would be positive when it is applied to a blood sample that does not contain HIV.

(b) Among the blood samples examined in the training program that provided positive ELISA results for HIV, what proportion actually contained HIV?

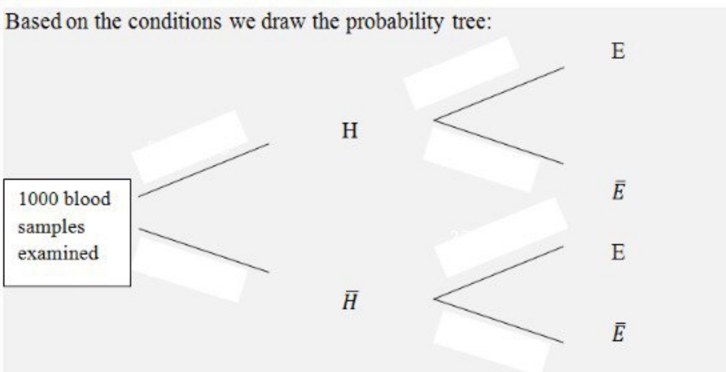
(c) When a blood sample yields a positive ELISA result, two more ELISAs are performed on the same blood sample. If at least one of the two additional ELISAs is positive, the blood sample is subjected to a more expensive and more accurate test to make a definitive determination of whether HIV is present in the sample. Repeated ELISAs on the same sample are generally assumed to be independent. Under the assumption of independence, what is the probability that a new blood sample that comes into the lab will be subjected to the more expensive test if that sample does not contain HIV?

Solution

Antistress hint: The problem looks quite scary, but actually the most difficult job here is to …read the whole long text of a problem! This problem is very typical for AP exam. Since this course is an elementary one, the only way to make the exam look more difficult is to complicate the tasks with additional useless information making the text longer. So, what the problem says? First of all ELISA is just a name of a medical test. Relax, you don't need to care about what it is and why it is called ELISA! (P.S. if you are interested, Google it!)

Let E be the event "positve ELISA". Let H be the event "blood sample contains HIV".

Based on the conditions we draw the probability tree:



(a) $P(E|\bar{H}) = \frac{37}{500} = \mathbf{0.074}$ [f]

(b) We can apply Bayes' rule:

$$P(H|E) = \frac{(P(E|H)P(H))}{(P(E|H)P(H) + P(E|\bar{H})P(\bar{H}))} =$$

$$= \frac{\frac{489}{500}\frac{500}{1000}}{\frac{489}{500}\frac{500}{1000} + \frac{37}{500}\frac{500}{1000}} = \frac{489}{489+37} = 0.9297.$$

Alternatively, we can use the definition of conditional probability: $P(H|E) = \frac{P(H \cap E)}{P(E)}$. As it is seen from the tree,

$$P(H \cap E) = \frac{489}{500}\frac{500}{1000} \text{ and } P(E) = \frac{489}{500}\frac{500}{1000} + \frac{37}{500}\frac{500}{1000}$$

Thus, $P(H|E) = 0.9297$.

Finally, you can explain your answer in the following way: a total of $489+37 = 526$ blood samples resulted in a positive ELISA. Of these, 489 samples actually contained HIV. Therefore the proportion of samples that resulted in a positive ELISA that actually contained HIV is: $489/526 = \mathbf{0.9297}$

Let E_1, E_2 and E_3 represent events of positive ELISA on 1st, 2nd and 3rd trial correspondingly.

As calculated in (a) for each trial $P(E_i|H) = 0.074$

The sample will be subjected to a more expensive test in three cases: $E_1 \cap E_2 \cap E_3$, or $E_1 \cap \bar{E}_2 \cap E_3$, or $E_1 \cap E_2 \cap \bar{E}_3$(these are the three elementary outcomes of the event "subjected to a more expensive test").

Thus,

Thus, $P(\text{"subjected to the more expensive test"}|\bar{H}) =$

$$= P(E_1 \cap E_2 \cap E_3|\bar{H}) + P(E_1 \cap \bar{E}_2 \cap E_3|\bar{H}) + P(E_1 \cap E_2 \cap \bar{E}_3|\bar{H})$$

As calculated in (a) for each trial $P(E_i|\bar{H}) = 0.074$. We are also told that the tree results are independent.

Therefore, $P(\text{“subjected to the more expensive test”}|\bar{H}) =$

$$= 0.074 \times 0.074 \times 0.074 +$$
$$+ 0.074 \times (1 - 0.074) \times 0.074 + 0.074 \times 0.074 \times (1 - 0.074) \approx \mathbf{0.0105}^{g}$$

Alternatively

$P(\text{“subjected to the more expensive test”}|\overline{HIV}) =$

$$= P(1^{\text{st}} \text{ ELISA positive and not both the } 2^{\text{nd}} \text{ and } 3^{\text{rd}} \text{ are negative}) =$$
$$= (0.074) \cdot (1 - 0.92622^2) \approx 0.0105$$

[f] Notice that we've estimated probability as the relative frequency of a positive ELISA among all the blood samples that does not have HIV.

[g] Notice that we've estimated probability as the sum of probabilities of its elementary outcomes. Probability of intersection of events equals the product of events' probabilities since they are known to be independent.

[g] Notice that we've estimated probability as the sum of probabilities of its elementary outcomes. Probability of intersection of events equals the product of events' probabilities since they are known to be independent.