

Discrete Probability Review

Lizzy Huang
Duke University

Week 1

The purpose of **Statistics with R** Specialization is to focus on implementing R code and using existing R functions / packages to solve statistical problems. However, we realize it is impossible to ignore the mathematical languages while we try to deliver the materials using lecture videos and R commands. In the previous 3 courses, we have minimized the use of mathematical formulas, due to the fact that concepts are relatively simple, so basic visualization and statistics tables will serve the purpose of solving these problems. In *Bayesian Statistics*, our goal is not to provide you the R functions and packages as blackboxes, but to explain in more mathematical details why Bayesian method works and how Bayesian approach differs from the frequentist approach you have seen previously. Therefore, mathematical notations and formulas are inevitable in explaining these more complex concepts. **However, the course expectation is still the same, that learners can perform most of the assignments using R functions and packages after watching the videos and reading the supplementary materials provided.** Hence, this **Review** aims to serve as a bridge to connect you from the concepts discussed in the previous 3 courses, to the corresponding mathematical notations and formulas you will see in the upcoming videos.

This **Review** is a summary of basic probability theories, not an introductory lecture note. That is to say, we do expect those who use this file have already gained some understanding of statistics concepts we have covered in the previous 3 courses and are familiar with basic probability terminology. This review cannot and should not be used as replacement of lectures or textbooks that teach you basic probability theories. If you have trouble with mathematical notations or have questions about terminology that we use here, please review the previous 3 courses in the specialization before you start this course.

Trial

A [trial](#), can be thought of as an experiment, or an attempt to do something. The content of a trial varies according to the context. For example, when we considered the probability of getting heads by flipping one coin, a trial would be “flipping a coin once”. But if we wanted to calculate the probability of getting two heads in a row by flipping coins, a trial would be “flipping a coin twice”.

Frequentist Definition of Probability

Now we have defined trial, we can talk about counting the outcomes we get from our trials. This leads to the idea of [probability under the frequentist definition](#). From Course 1: **Introduction to Probability and Data**, we provided the frequentist definition of probability as:

The probability of an outcome is the proportion of times the outcome would occur if we have repeated a trial an infinite number of times.

While the frequentist definition of probability requires us to conduct the same trial (or experiment) an infinite number of times, the [Bayesian interpretation of probability](#) does not depend on the number of times we conduct the trial. Instead, it interprets probability as

A subjective degree of belief

This has reduced the “workload” of considering the limiting case when a trial is repeated an infinite number of times. However, it will introduce personal bias because this belief is subjective. This is why in this week, we focus on how to update our belief, using [Bayes’ Rule](#).

Sample Space

A **sample space** is a collection of all the possible outcomes of a trial. For example, if you flipped a coin twice, all the possible outcomes would be

$$\{HH, HT, TH, TT\}.$$

(We use H to represent getting 1 head, and T to represent getting 1 tail. The order of getting heads and tails matters, so HT and TH are different outcomes.)

We recommend writing down the sample space for simple probability problems, so that one will not miss any possible results when calculating probability. We have seen an example in the first Course **Introduction to Probability and Data**, Week 4 Video **Binomial Distribution** when Professor Mine Rundel showed us how to enumerate all the possible outcomes in the Milgram's experience example. In that video, we counted the number of all possible outcomes when 1 out of 4 "teachers" chose to give a shot, and came up with the constant for the Binomial distribution.

Event

In general, one will not ask the question "What is the probability of a sample space?" More usual, the question would be "What is the probability of an event?". An **event** is a smaller collection of the outcomes of the sample space, or the sample space itself. For example, in the "flipping a coin twice" example, we could ask: what is the probability of getting at least 1 head. Or we could ask: what is the probability of getting exactly 1 tail.

An event can be the sample space itself (including all possible outcomes), or it can be an empty set (including no outcome).

About Event Notation

One may use text description for an event, however, this is not that efficient. Therefore, we usually use capital alphabet letter to denote an event. Say

Let A be the event "getting at least 1 head if we flip a coin twice".

Let B be the event "getting exactly 1 tail if we flip a coin twice".

Very rare, but it happens, that we may use letters other than capital alphabet letter to denote events. These happen when we run out of all 26 letters, or when we try not to use the same letter to mean 2 different things. One can find out the meaning of notations from the context.

Union and Intersection

A **union** of event A and event B is the collection of outcomes that are either in A or B . These outcomes can be in both events A and B . We use A or B , or sometimes, $A \cup B$ to denote the union. In the "flipping a coin twice" example, the union of A and B has outcomes:

$$A \text{ or } B = \{HH, HT, TH\}.$$

An **intersection** of event A and event B is the collection of outcomes that are in both A and B . We use A and B , or sometimes, $A \cap B$ to denote the intersection. In the "flipping a coin twice" example, the intersection of A and B has outcomes:

$$A \text{ and } B = \{HT, TH\}.$$

Disjoint Event

Two events are said to be **disjoint** if they do not share any common outcomes. That means, if events A and B are disjoint, then the intersection A and B is empty. Apparently, the events A and B in the previous “flipping a coin twice” example share common outcomes HT , TH , so they are not disjoint.

Complementary Event

A **complementary event** of an event A is the collection of outcomes that are *not* included in event A . It is obvious that an event and its complementary event are disjoint, and the union of them is the entire sample space. We denote the complementary event of an event A as A^c .

About Probability Notation

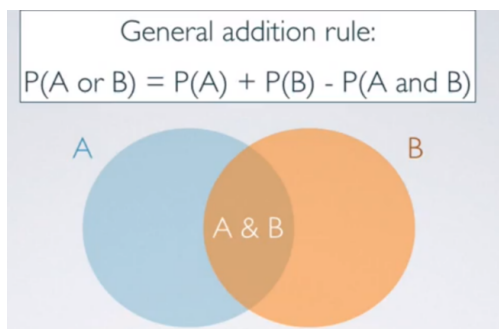
Now we have discussed the outcomes and their collections, we can talk about probability for these outcomes or their collections to happen. Before we review the calculation of probability, we want to clarify the mathematical notation we usually use (at least in this course) for the probability of an event. The probability of an event A is usually denoted as

$$P(A).$$

We use capital letter P for the notation of the probability of events. To guarantee no confusion in mathematical meaning, we in general will avoid using P to denote an event, as you can see, $P(P)$ is a bad notation.

General Addition Rule

From the Venn diagram below,



we can derive the **general addition rule** when we try to add the probabilities of 2 events A and B

$$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B). \quad (1)$$

This formula is useful in calculating $P(A \text{ or } B)$, when it is easier to calculate $P(A)$, $P(B)$, and $P(A \text{ and } B)$ instead of $P(A \text{ or } B)$ directly. We have gone over examples in the first course, so we will not repeat them here.

When event A and event B are disjoint, we can simplify Equation (1) to

$$P(A \text{ or } B) = P(A) + P(B) - \cancel{P(A \text{ and } B)} \rightarrow 0$$

For an event A and its complementary event A^c , since these two events are disjoint by definition, and the union of them is the entire sample space, we can further simplify the *general addition rule* to be

$$\cancel{P(A \text{ or } A^c)}^1 = P(A) + P(A^c) - \cancel{P(A \text{ and } A^c)}^0 \implies P(A) = 1 - P(A^c).$$

This formula is useful when calculating $P(A^c)$ is easier than calculating $P(A)$ directly.

Binomial Distribution

To discuss binomial distribution, we first need to talk about its building blocks, the Bernoulli trial.

A **Bernoulli trial** is a trial which has only 2 possible outcomes. For example, when flipping a coin once, we can only have H or T . When taking a course, one student can either “pass” or “fail”. We usually call one outcome the **success** (the one that we care more), and the other the **failure**. If we assign probability p for the **success** to occur (also known as the **success rate**), then the **failure** must have probability $1 - p$, for some p between 0 and 1.

The **Binomial distribution** gives the **probability** (not the trial anymore) of having exactly k successes in n independent Bernoulli trials with probability of success p .

$$P(\text{exactly } k \text{ successes in } n \text{ independent Bernoulli trials with probability of success } p) = \binom{n}{k} p^k (1 - p)^{n-k}. \quad (2)$$

Here the notation $\binom{n}{k}$ is called “ n chooses k ”, whose value is given by the formula $\binom{n}{k} = \frac{n!}{k!(n-k)!}$. And $k! = 1 \times 2 \times \cdots \times k$, the first k whole numbers multiplying together.

Example 1

Let us walk through an example combining all the concepts we have reviewed so far.

Suppose in a bag of M&M’s with 5 different colors: **red**, **green**, **yellow**, **blue**, and **orange**. Suppose the general percentage of yellow M&M’s in a bag is 10%, i.e., the probability of getting a yellow M&M is 0.1. What would be the probability of getting 1 or more yellow M&M in a bag of 5 M&M’s?

The answer is not $\frac{1}{5}$, since we never assume that all colors of M&M’s are distributed evenly in a bag. It is not 0.1 either, since we ask for **more than 1** yellow M&M. To solve this problem, we need to use the Binomial distribution. We view the outcome of a Bernoulli trial as: getting yellow M&M ($p = 0.1$), not getting yellow M&M. And we have 5 M&M’s in a bag, so there are 5 independent Bernoulli trials. The formula (2) only gives us the probability of getting **exactly** k successes, not the same as what the problem asks for. So to solve this problem, we need to leverage the **general addition rule**. Let us go through the solution.

The event “having 1 or more yellow M&M’s in a bag of 5” can be broken down into the following 5 events: having exactly 1 yellow M&M, having exactly 2 yellow M&M’s, having exactly 3 yellow M&M’s, having exactly 4 yellow M&M’s, and having exactly 5 yellow M&M’s. Since all these events are disjoint, the *general addition rule* gives us

$$\begin{aligned} &P(1 \text{ or more yellow M\&M's}) \\ &= P(1 \text{ yellow M\&M}) + P(2 \text{ yellow M\&M's}) + P(3 \text{ yellow M\&M's}) + P(4 \text{ yellow M\&M's}) + P(5 \text{ yellow M\&M's}) \end{aligned} \quad (3)$$

Now we can use the formula (2) of Binomial distribution to calculate the 5 probabilities. Here, the number of “successes” k is 1, 2, 3, 4, and 5, for each case, and the total number of “trials” n is 5. The “probability of success” $p = 0.1$. We further formulate the problem using math

$$\begin{aligned}
 &P(\text{having 1 or more yellow M\&M's}) \\
 &= P(1 \text{ yellow M\&M}) + P(2 \text{ yellow M\&M's}) + P(3 \text{ yellow M\&M's}) + P(4 \text{ yellow M\&M's}) + P(5 \text{ yellow M\&M's}) \\
 &= \binom{5}{1}(0.1)^1(1-0.1)^{5-1} + \binom{5}{2}(0.1)^2(1-0.1)^{5-2} + \binom{5}{3}(0.1)^3(1-0.1)^{5-3} \\
 &\quad + \binom{5}{4}(0.1)^4(1-0.1)^{5-4} + \binom{5}{5}(0.1)^5(1-0.1)^{5-5}
 \end{aligned} \tag{4}$$

If we were to calculate all 5 probabilities using the Binomial distribution formula, we would need to compute many factorials. We would like to seek a faster way to get to the answer. In this case, *complementary event* can be of assistance. Recall that we can calculate $P(A)$ using $1 - P(A^c)$. If the complementary event A^c is simpler event, our calculation will be less. In this M&M's example, the complementary event of “getting 1 or more yellow M&M's” is “getting exactly 0 yellow M&M”. We can easily get its probability

$$P(0 \text{ yellow M\&M}) = \binom{5}{0}(0.1)^0(1-0.1)^{5-0} = (0.9)^5$$

Therefore, the probability of getting 1 or more yellow M&M's is

$$P(1 \text{ or more yellow M\&M's}) = 1 - P(0 \text{ yellow M\&M}) = 1 - (0.9)^5 \approx 0.41.$$

R Code

We presented the above mathematical solution, in order to show how to break down a complicated problem and solve it using multiple probability concepts combined together. Once you have had a good idea how to approach the problem, you may use R functions to help you for the rest of the computation (including computing factorials). The Binomial distribution is given as `dbinom` function. When you type in `?dbinom` in the R Console, you will see the help document listing all the functions associated with the Binomial distribution in R. The above example can be solved using either

```
prob = dbinom(1, 5, 0.1) + dbinom(2, 5, 0.1) + dbinom(3, 5, 0.1) +
       dbinom(4, 5, 0.1) + dbinom(5, 5, 0.1)
prob
```

```
## [1] 0.40951
```

if you prefer the complicated route, or

```
prob = 1 - dbinom(0, 5, 0.1)
prob
```

```
## [1] 0.40951
```

if you can leverage the complementary event to help you. We get same results from each approach. No matter which approach you choose, it is very important to be able to sort out the steps and formulate them using mathematical definitions.

Random Variable (new)

This course mainly focuses on how to apply R functions to obtain results. However, we also hope to convey the mathematical ideas so that these functions and code will not seem to be blackboxes to you. To communicate mathematics, it is inevitable that we need abstract notations for concepts and formulas, to reduce the load of keeping track of long sentences and to make these concepts and formulas more applicable in general cases.

We start with the more abstract concept, *random variable*. Imagine if we have a bag of 1000 M&M's, and we want to calculate the probability of getting 0, 1, 2, 3, until 1000 yellow M&M's, it is not an ideal way to really start the following list

$$\begin{aligned}P(\text{exactly 0 yellow M\&M}) &= \binom{1000}{0} (0.1)^0 (1 - 0.1)^{1000-0} \\P(\text{exactly 1 yellow M\&M}) &= \binom{1000}{1} (0.1)^1 (1 - 0.1)^{1000-1} \\P(\text{exactly 2 yellow M\&M's}) &= \binom{1000}{2} (0.1)^2 (1 - 0.1)^{1000-2} \\&\vdots \\P(\text{exactly 1000 yellow M\&M}) &= \binom{1000}{1000} (0.1)^{1000} (1 - 0.1)^{1000-1000}\end{aligned}$$

This is why we abstract the “number of yellow M&M's” to be k , and use one formula below to include all the situations describe by the above 1001 equations.

$$P(\text{exactly } k \text{ yellow M\&M's}) = \binom{1000}{k} (0.1)^k (1 - 0.1)^{1000-k}, \quad k = 0, 1, 2, \dots, 1000.$$

This is not generalized enough. What if we ask for any bag of arbitrary number of M&M's and any percentage of yellow M&M's, and we still want to list all the probabilities? We know the answer, the Binomial definition formula (2) tells us we can further abstract the total number as n , and the probability of “success” to be p . Then the formula has become

$$P(\text{exactly } k \text{ yellow M\&M's}) = \binom{n}{k} p^k (1 - p)^{n-k}.$$

But this is not concise enough, since we still need to use words to describe the events that we care “exactly k yellow M&M's”. We have seen that, for a given situation when n and p are known, the only varying number is the value of k . Hence, to make sure readers understand the problem we are solving (it's about yellow M&M's, not the number of heads in coin flips), but also emphasize the varying value, we further use [random variable](#) to help denoting probability events.

In the M&M's example, we can denote X , the **random variable**, to be the number of yellow M&M's we get in a bag of n M&M's with yellow M&M percentage p . Then the probability of getting exactly k yellow M&M's is

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$$

Here, for each k value, $X = k$ is an event. Translated by to English, this means “the number of yellow M&M's (X) equals to k ”. Since $X = k$ denotes an event for each k , we follow the convention and use P as the probability notation.

Benefits of Using Random Variable

As we have discussed, random variable notation helps us to save words and time. Moreover, there are sound mathematical reasons to use the notation $X = k$ to denote the event of interest. One good property about random variables is, $X = k$ for different k 's are disjoint events. That means, $X = 1$ and $X = 2$ will never share any common outcomes. And all " $X = k$ " events together cover all outcomes in the sample space. Therefore, the *general addition rule* is pretty simple here. For any random variable X , the following is always true

$$\sum_{\text{all } k} P(X = k) = 1.$$

However, as we discussed in Course 1: *Introduction to Probability and Data*, for any 2 events A and B , the followings are in general **not true**

$$P(A) + P(B) = 1 \text{ (incorrect!)}, \quad \text{or} \quad P(A) + P(B) \leq 1 \text{ (incorrect!).}$$

More Probability Notations and Parameters

Once a Binomial probability problem is given, the total number of trials n , and the probability of success p will be fixed. The only value that can vary according to the question is k . Looking back to the Binomial formula (2), we see that it is actually a function of k . We denote this function as $p(k)$. Its input (or independent variable) is k , and its output is the probability

$$p(k) = P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}. \quad (5)$$

Whenever we denote a probability mass function (for discrete random variable) or a probability distribution function (for continuous random variable), we use lower case p (some independent variable) as our notation.

The total number of trials n and the probability of success p are called **parameters** of the Binomial distribution. Although they are fixed and will be provided by the problem, somehow they can take different values. So the values of $p(k)$ from Equation (5) also depend on n and p . Therefore, Equation (5) actually gives a **family of functions**: each n and p combined together correspond to a specific Binomial distribution. Within this specific Binomial distribution, changing values of k provides different probabilities for each event: k successes out of the total trials. Therefore, sometimes we also denote the Binomial distribution as

$$B(n, p)$$

Question 1: we use lower case p to denote probability mass/distribution functions, but inside the Binomial distribution, we also have parameter p . Isn't the 2 inconsistent?

Answer: it is, in the case when the general notation p is needed if the problem does not provide its value. That is why we sometimes just use $P(X = k)$ instead of $p(k)$, or we may use $f(k) = \binom{n}{k} p^k (1 - p)^{n-k}$ to avoid ambiguity.

Question 2: can parameters such as n or p have their probability distributions? What about the notations in this case?

Answer: yes they can. In this case, we still stick with lower case p as the probability function notation. However, $p(p)$ is again a bad notation. Therefore, in Week 2, whenever we talk about the probability distribution of the "probability of success" in the Binomial distribution, we use the Greek letter π as the probability function notation: $\pi(p)$.

Conditional Probability

Bayesian statistics aims to answer the following question:

Given the data we observed, what is the probability for some event of interest happens?

To “translate” this question into mathematics, we need [conditional probability](#). The probability of event A given event B , or “conditioning on event B ” is denoted as $P(A | B)$. With this notation, we can formulate the question above as

What is $P(\text{event of interest} | \text{data observed})$?

$P(A | B)$ **in general is not equal to** $P(A)$. A simple example to keep in mind is:

Suppose we flip a fair coin once. If we do not know the result, the probability of getting a head $P(\text{head}) = \frac{1}{2}$. However, if we already know that the coin flip gives a head, the probability of getting a head given that we know it is a head is $P(\text{head} | \text{head}) = 1$.

General Product Rule

The idea of conditional probability can be viewed as we break down a probability problem step by step. The first step, the “condition” we have, the second step, the event “based on this condition”. These 2 steps as a whole gives us the probability of both the event and the condition happen together. We formulate this idea using the [general product rule](#):

$$P(\text{event and condition together}) = P(\text{event} | \text{condition}) \times P(\text{condition}),$$

or using letter notations to replace events and conditions, we have

$$P(A \text{ and } B) = P(A | B) \times P(B).$$

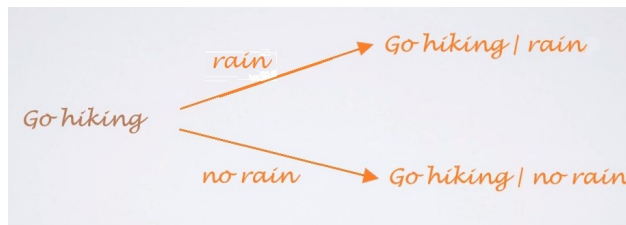
This provides a formula to back solve the *conditional probability*:

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

Law of Total Probability (new)

With the idea of conditional probability in mind, we would like to ask, how can we calculate the probability of an event A through breaking it down into steps?

Imagine that you are interested in knowing the probability of your friend going to hike with you tomorrow. Your friend says, “If it rains tomorrow, there would be 10% of the chance that I would come. If it does not rain tomorrow, the chance would be higher, 70%.” You know for tomorrow’s weather, it either rains, or does not rain. These 2 cannot happen together (so they are disjoint). When you try to calculate the total probability that your friend would go hiking, you break it into 2 paths: “rain” and “no rain”.



$$P(\text{your friend comes}) = P(\text{your friend comes and it rains}) + P(\text{your friend comes and it does not rain}).$$

We further break the 2 probabilities using the *general product rule*.

In general, the **Law of total probability** tells us, if the we have a finite (or countably infinite) set of disjoint events $\{B_1, B_2, \dots\}$ that partition a sample space (in the example, sample space = tomorrow’s weather), we can calculate the probability of any event A as

$$P(A) = P(A | B_1)P(B_1) + P(A | B_2)P(B_2) + \dots = \sum_{\text{all } B_i} P(A | B_i)P(B_i).$$

When we partition the sample space into just 2 parts, an event B and its complementary event B^c . The *Law of total probability* can be simplified as

$$P(A) = P(A | B)P(B) + P(A | B^c)P(B^c).$$

We will present an example later to demonstrate how to use these concepts to solve problems.

Independence

The last concept that was introduced in the course of **Introduction of Probability and Data** is *independence*. Two events are **independent** if knowing the outcome of one event provides no useful information about the outcome of another event. Intuitively speaking, this is to say, the first step (the event as a “condition”) provides no useful information about the second step (the event of interest). Mathematically, we have

$$P(\text{event} | \text{condition}) = P(\text{event}), \quad \text{or,} \quad P(A | B) = P(A).$$

If 2 events are independent, the *general product rule* becomes

$$P(A \text{ and } B) = P(A | B) \times P(B) = P(A) \times P(B).$$

Independence is mutual, so if A is independent of B , then B is also independent of A , as long as $P(A)$ and $P(B)$ are non-zero. That is,

$$P(A | B) = P(A) \implies P(B | A) = P(B).$$

Independence in Conditional Probability

We have talked about viewing conditional probability as the second step “conditioning on” the first step. But what if a problem has more than 2 steps? Of course, we will have the corresponding conditional probability formula on multiple conditions. However, we do not want to scare you with the heavy notations and equal signs. Instead, we only discuss a common situation **that we will see in this course**, to leverage independence within conditional probability.

Suppose event A and event B are independent, after conditioning on event C , they are still independent. That is

$$P(A \text{ and } B \mid C) = P(A \mid C) \times P(B \mid C). \quad (6)$$

In general, independence is not guaranteed.

The “independent” concept is different from “disjoint”. More discussion is provided by **Week 3 Video: (Spotlight) Disjoint vs. Independent** in Course 1 **Introduction to Probability and Data**.

Example 2

It is time for us to bring all the concepts together into an example.

In the 1980s, the U.S. Military provided the enzyme-linked immunosorbent assay (ELISA) test to test for human immunodeficiency virus (HIV) among recruits. The true positive (sensitivity) of the test was around 93%, while the true negative (specificity) of the test was around 99%. It was estimated that 1.48 / 1000 adult Americans were HIV positive. Suppose the result of each ELISA test is independent with each other.

Question 1:

What is the probability of a recruit who has HIV getting 1 positive and 1 negative if they got 2 ELISA tests?

To answer this question, we first need to formulate the information we have in the question using mathematics. This step will help us to better understand what the question wants us to answer and what tools we have to answer this question.

Step 1: “translate” information into math:

From the question, we have the following information:

true positive of ELISA is 93%	$\implies P(\text{ELISA} + \mid \text{has HIV}) = 0.93$
true negative of ELISA is 99%	$\implies P(\text{ELISA} - \mid \text{no HIV}) = 0.99$
1.48 / 1000 adult Americans were HIV positive	$\implies P(\text{has HIV}) = 0.00148$
probability getting 1 positive and 1 negative given HIV	$\implies P(1 \text{ ELISA} + \text{ and } 1 \text{ ELISA} - \mid \text{has HIV})$

(7)

Step 2: use information we have and probability formulas to answer the question:

The last probability in (7) is the answer we need. We have already had the probability $P(\text{ELISA} + \mid \text{has HIV})$. Now the question asks for 1 positive and 1 negative ELISA results, given the recruit has HIV.

This process has 2 independent Bernoulli trials, with 1 “success” (test positive when having HIV), with probability $p = P(\text{ELISA} + \mid \text{has HIV}) = 0.93$. So the probability can be simply calculated using Binomial distribution.

$$P(1 \text{ ELISA} + \text{ and } 1 \text{ ELISA} - \mid \text{has HIV}) = \binom{2}{1} p^1 (1-p)^{2-1} = 2 \times 0.93 \times (1 - 0.93) = 0.1302.$$

This is an application of the formula (6).

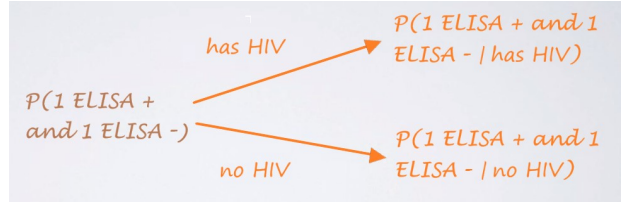
Question 2:

What is the probability of a recruit getting 1 positive and 1 negative ELISA results?

Question 2 is different from Question 1 where we no longer have the conditional information that this recruit has HIV. We still first formulate this question as finding

$$P(1 \text{ ELISA} + \text{ and } 1 \text{ ELISA} -).$$

We have already got the probability $P(1 \text{ ELISA} + \text{ and } 1 \text{ ELISA} - \mid \text{has HIV})$. That this recruit has HIV is one possible condition of getting the ELISA test results. Another possible condition is that this recruit does not HIV. That means, we can break the question into 2 paths using the probability tree.



This time, we use the *Law of total probability* to calculate the total probability $P(1 \text{ ELISA} + \text{ and } 1 \text{ ELISA} -)$.

$$P(1 \text{ ELISA} + \text{ and } 1 \text{ ELISA} -) \tag{8}$$

$$= P(1 \text{ ELISA} + \text{ and } 1 \text{ ELISA} - \mid \text{has HIV})P(\text{has HIV}) + P(1 \text{ ELISA} + \text{ and } 1 \text{ ELISA} - \mid \text{no HIV})P(\text{no HIV}). \tag{9}$$

We only need to calculate the numbers of the terms $P(1 \text{ ELISA} + \text{ and } 1 \text{ ELISA} - \mid \text{no HIV})$ and $P(\text{no HIV})$.

“Having no HIV” is the complementary event of “having HIV”. Therefore, the probability of having no HIV is

$$P(\text{no HIV}) = 1 - P(\text{has HIV}) = 1 - 0.00148 = 0.99852.$$

“Having 1 positive and 1 negative ELISA results” given the recruit has no HIV, can be viewed as another 2 independent “Bernoulli trials” with 1 “success” (test negative when having no HIV) and probability of success $p = P(\text{ELISA} - \mid \text{no HIV}) = 0.99$. We can calculate this probability using another Binomial distribution

$$P(1 \text{ ELISA} + \text{ and } 1 \text{ ELISA} - \mid \text{no HIV}) = \binom{2}{1} 0.99^1 (1 - 0.99)^1 = 0.0198.$$

Combining all the terms together, we get

$$\begin{aligned}
& P(1 \text{ ELISA } + \text{ and } 1 \text{ ELISA } -) \\
&= P(1 \text{ ELISA } + \text{ and } 1 \text{ ELISA } - \mid \text{has HIV})P(\text{has HIV}) + P(1 \text{ ELISA } + \text{ and } 1 \text{ ELISA } - \mid \text{no HIV})P(\text{no HIV}) \\
&= (0.1302) \times (0.00148) + (0.0198) \times (0.99852) \\
&\approx 0.01996.
\end{aligned}$$

R Code

If you are not a big fan of doing mathematical arithmetic, you may solve these questions using R. However, we highly suggest you formulate everything using mathematical notations and write down clear steps, before implementing any of them using R functions.

```

# Question 1
prob = dbinom(1, 2, 0.93)
prob

## [1] 0.1302

# Question 2
prob = dbinom(1, 2, 0.93) * (1.48 / 1000) + dbinom(1, 2, 0.99) * (1 - 1.48 / 1000)
prob

## [1] 0.01996339

```

If you do not know whether a recruit has HIV or not, getting only around 2% chance after one ELISA positive and one ELISA negative is not that scary, even with high percentages of true positive and true negative.

In this week, we will derive the [Bayes' rule](#), which can help us to calculate the probability of having HIV after seeing the ELISA results, using. In the lectures, we use **sequential updating**, another method rather than viewing Bernoulli trials together as a Binomial process, to get the same result. Finally, we will illustrate the difference between frequentist inference and Bayesian inference, using yellow M&M's and other medical examples.