**Lesson 5**

**Researches about theory (R)**

6\_R. Think and explain in your own words what is the role that probability plays in Statistics and the relation between the observed distribution and frequencies their "theoretical" counterparts. Do some practical examples where you explain how the concepts of an abstract probability space relate to more "concrete" and "real-world" objects when doing statistics.

To explain this concept, I want to make an example first and then I will try to explain it in more clear way. Let’s take for example a class of 100 students (as always students from Cybersecurity course). These students took the Cryptography exam, and we are observing the results. We don’t have all the results, but we have only a partial view of them, for example 20 results out of 100. These 20 results are factual numbers, in the sense that they cannot be contested as “false”, so we know for sure those 20 results. We can draw a distribution of the 20 results considering that the total space of the results is 20. We can define some grades classes, like “insufficient”, “low”, “medium”, “high” and “cum laude”.

|  |  |
| --- | --- |
| Grades class | Meaning |
| Insufficient | < 18 |
| Low | 18 <= grade < 22 |
| Medium | 22 <= grade < 26 |
| High | 26 <= grade <= 30 |
| Cum Laude | 30 cum laude |

Considering these classes, we can observe the distribution of the 20 students in the Cryptography exam.

Chart, waterfall chart

Description automatically generated

So, we can observe that 8 students took an insufficient grade, 5 students took a low grade, 3 students took a medium grade, 3 students took a high grade and only 1 student took a grade cum laude. These grades or this distribution cannot be contested because they are factual. They are observations from the real world.

What about the other 80 students? We don’t know which grades they took. For sure we could try to guess those grades. But which probabilities we can assign to each grades class? We can use the grades of the 20 students we know!

|  |  |
| --- | --- |
| Grades classes | Probability |
| Insufficient | 8/20 |
| Low | 5/20 |
| Medium | 3/20 |
| High | 3/20 |
| Cum Laude | 1/20 |

Using these probabilities, we can try to infer the total distribution, so the most probable distribution would be:

Chart, bar chart

Description automatically generated

This is the distribution showing the grades of the 100 students’ results based on the 20 results grades we know. But there is one important thing to say now: these are not facts, we are just trying to understand which is the most likely distribution of all grades, but of course now we can contest this distribution.

There are a lot of distributions (of all the 100 students) that can be generated, the most likely is this one, but we don’t know if this is the real one or not.

When we have some evidence, which usually is a sample (statistical) from which we can draw our dataset, we can do for instance a statistical distribution. We can refer to this statistical distribution as ‘empirical’ because empirical means that it is based on some facts, observations (From Oxford Dictionary: “based on, concerned with, or verifiable by observation or experience rather than theory or pure logic.”). In Statistical Inference, that is different from Descriptive (in Descriptive we only care about describing the observations we observed in a good way), we are imaging there is a theoretical distribution rising from the observation. We can refer to this as a ‘model’.

7\_R. Explain the Bayes Theorem and its key role in statistical induction. Describe the different paradigms that can be found within statistical inference (such as "bayesian", "frequentist" [Fisher, Neyman]).

From Wikipedia: “In probability theory and statistics, Bayes' theorem (alternatively Bayes' law or Bayes' rule; recently Bayes–Price theorem), named after Thomas Bayes, describes the probability of an event, based on prior knowledge of conditions that might be related to the event. For example, if the risk of developing health problems is known to increase with age, Bayes' theorem allows the risk to an individual of a known age to be assessed more accurately (by conditioning it on their age) than simply assuming that the individual is typical of the population.

One of the many applications of Bayes' theorem is Bayesian inference, a particular approach to statistical inference. When applied, the probabilities involved in the theorem may have different probability interpretations. With Bayesian probability interpretation, the theorem expresses how a degree of belief, expressed as a probability, should rationally change to account for the availability of related evidence. Bayesian inference is fundamental to Bayesian statistics.”

This is the formula of Bayes Theorem:



Where:

* A and B are events and P(B) != 0.
* P(A|B) is a conditional probability: the probability of event A occurring given that B is true. It is also called the posterior probability of A given B.
* P(B|A) is also a conditional probability: the probability of event B occurring given that A is true. It can also be interpreted as the likelihood of A given a fixed B because of P(B|A) = L(A|B).
* P(A) and P(B) are the probabilities of observing A and B respectively without any given conditions; they are known as the marginal probability or prior probability.
* A and B must be different events.

We just saw an example of how this theorem is used in Statistics: given the 20 results of the students in the Cryptography exam, which is the most likely total (100 results) distribution that gave birth to those 20 results? Using the Bayes theorem, we can observe how much a model is likely to be that one.

The interpretation of Bayes' rule depends on the interpretation of probability ascribed to the terms. The two main interpretations are described below.

In the Bayesian (or epistemological) interpretation, probability measures a "degree of belief". Bayes' theorem links the degree of belief in a proposition before and after accounting for evidence. For example, suppose it is believed with 50% certainty that a coin is twice as likely to land heads than tails. If the coin is flipped a number of times and the outcomes observed, that degree of belief will probably rise or fall, but might even remain the same, depending on the results. For proposition A and evidence B,

* P (A), the prior, is the initial degree of belief in A.
* P (A|B), the posterior, is the degree of belief after incorporating news that B is true.
* the quotient P(B|A)/P(B)represents the support B provides for A.

In the frequentist interpretation, probability measures a "proportion of outcomes". For example, suppose an experiment is performed many times. P(A) is the proportion of outcomes with property A (the prior) and P(B) is the proportion with property B. P(B|A) is the proportion of outcomes with property B out of outcomes with property A, and P(A|B) is the proportion of those with A out of those with B (the posterior).

**Applications / Practice (A) [work on this at least 30' a day, all days]**

7\_A. Given 2 variables from a csv compute and represent the statistical regression lines (X to Y and vice versa) and the scatterplot.  
Optionally, also represent the histograms on the "sides" of the chart (one could be draw vertically and the other one horizontally, in the position that you prefer).  
[Remember that all our charts must always be done within "dynamic viewports" (movable/resizable rectangles). No third-party libraries, to ensure ownership of creative process. May choose the language you prefer.].

WRITE HERE

**Researches about applications (RA)**

5\_RA. Do web research about the various methods to generate, from a Uniform ([0,1)), all the most important random variables (discrete and continuous). Collect all source code you think might be useful code of such algorithms (keep credits and attributions wherever applicable), as they will be useful for our next simulations. v

WRITE HERE

**References**