FINITE PROBABILITIES

Contents

Module

This chapter is part of the DP Module ("Dénombrement et Probabilités", that is, "Counting, Enumerating and Probabilities").

The prerequisites are: the LE Module ("Logic and Sets", seen during the seminar).

Motivation

Any observable phenomenon has an element of randomness. Sometimes, this element is only due to a lack of information. For example, if we toss a coin on a table, it should be possible to predict whether the result will be ***heads*** or ***tails***: if we could accurately know the initial position of the coin, the exact forces it has been subjected to, the air pressure, the table granularity, and so on, then it would be possible to compute the coin's exact movement and the final result. In such a situation, randomness comes from the difficulty to collect the required information: it is not reasonable to measure accurately so many physical parameters and to process them. Nevertheless, other phenomena are intrinsically random. The movement of a physical particle cannot be observed without disturbing the particle[1]. It is hence not possible to observe the trajectory. The latter is an intrinsic random phenomenon.

The purpose of probability theory is modeling random phenomena. It is a very wide theory but, to start with, we will restrict to ***finite*** probabilities. In this framework, a random experiment has only a ***finite*** number of possible outcomes. This includes the result of rolling a dice, drawing numbers at a lottery or picking cards from a deck. Focusing on finite cases enables one to study most of the concepts from the probability theory, while avoiding the additional complexity which would result from infinite sets' manipulation. Furthermore, many real-life situations are handled with finite probabilities. For example, determining the expectation of an algorithm's complexity requires hypotheses about the algorithm's inputs. These hypotheses can be done using ranged values for these inputs (time delay to connect to a server, etc.). The risk involved in choosing one path from a finite number of possible paths, when moving in a graph, can also be modeled in this framework. In the general case, probabilities obviously appear in machine learning, image analysis and statistics – this area should be studied by any engineer, especially in our times with the emergence of ***big data*** technologies.

Learning outcomes

The learning outcomes, which will be tested at the exams, are the following:

- model a random experiment using a probability language;
- compute probabilities in equiprobable situations, given conditional probabilities or an independence hypothesis;
- interpret the expectation and variance of a finite random variable and of any linear transformation of this variable.

To give more details:

1. model a random experiment using a suitable probability space: find the set of possible outcomes, the events, the probability function;
2. compute probabilities in case of equiprobability;
3. compute conditional probabilities (using a tree or not) in case of a finite sample set;
4. compute conditional probabilities using the formula of total probabilities;
5. interpret the independence of two events;

[1] Heisenberg's uncertainty principle

6. find the distribution, the expectation and the variance of a random variable;
7. express some random variables as a sum of simpler random variables (for example, a binomial variable is a sum of Bernoulli variables).

When reasoning about an exercise, keep in mind the two following points:

- a probability can be neither negative nor greater than 1, it is important to check that your final result is consistent
- mutually exclusive events are not always independent. Conversely, independent events are not always mutually exclusive.

1 Modeling a random experiment

1.1 Summary

Quantifying randomness consists in mapping each possible outcome of a random experiment to a proportion (a real number between 0 and 1). The proportion is the «chance» that this possible outcome is the final one, which will result from running the experiment. When you are faced with a random experiment, choosing the quantification is the ***modeling*** step. You have to choose a model which fits your experience. During this step, check that the quantification is consistent with the basic properties that any probability function must satisfy.

Fitting a random experiment to your experience can lead to define a random variable. Such a variable is a function which maps each possible outcome of the experiment to a real number. In many situations, after running the experiment, we observe only this real number instead of the exact outcome. The distribution of the random variable gives the probability of each possible value of the variable. This probability is the sum of the probabilities of all the outcomes leading to this numerical value. It is very important to be able to model a random experiment both in terms of outcomes and in terms of variable distribution: in some cases, you will have experience with the possible outcomes and their probabilities, while in other cases, you will only have experience with a random variable.

1.2 Exercises

Exercise 1.1
Let Ω be a set and $f : \mathcal{P}(\Omega) \longrightarrow \mathbb{R}$ a mapping satisfying the conditions[2]:

- $f(\Omega) = 1$
- $\forall A \in \mathcal{P}(\Omega), \quad f(A) \geqslant 0$
- $\forall (A, B) \in \mathcal{P}(\Omega)^2, \quad \text{if} \quad A \cap B = \emptyset \quad \text{then} \quad f(A \sqcup B) = f(A) + f(B)$

Show the following relations:

1. $f(\emptyset) = 0$;
2. $\forall A \in \mathcal{P}(\Omega)$, $0 \leqslant f(A) \leqslant 1$;
3. $f(\overline{A}) = 1 - f(A)$;
4. if $A \subset B$, $f(A) \leqslant f(B)$ and $f(B \backslash A) = f(B) - f(A)$;
5. for any A, $B \in \mathcal{P}(\Omega)$, $f(A \cup B) = f(A) + f(B) - f(A \cap B)$.

How would you call the 3-tuple $(\Omega, \mathcal{P}(\Omega), f)$? How would you call the function f?

[2]The Kolmogorov's axioms.

Exercise 1.2
For each of the the following experiments, display the sample set Ω and determine the probability of each possible outcome.

1. tossing a coin
2. rolling a fair dice
3. rolling two fair dice, each of them having a different color (then you can distinguish them)
4. drawing one marble from a box containing 5 red and 6 white marbles
5. drawing with replacement three marbles from a box containing 5 red and 6 white marbles
6. drawing simultaneously three marbles from a box containing 5 red and 6 white marbles
7. drawing successively three marbles from a box containing 5 red and 6 white marbles.

Exercise 1.3
At the time when mathematicians were wondering how to build the probability theory, H. Poincaré asked the following question: when rolling two indistinguishable dice, we can get the probability $\frac{11}{36}$ or $\frac{6}{21}$ for the event «getting at least one 6»! Explain his two ways of reasoning. Which one is correct?

Exercise 1.4
1. In a population of $N > 1$ persons, what is the smallest value of N to be sure that at least two persons have the same birthday?
2. What is the probability that at least two persons, from a size $N > 1$ population, have the same birthday?[3]

2 Equiprobability

2.1 Summary

Equiprobability is a property of the probability function which states that each possible outcome has the same probability. The distribution is uniform and the common probability is the inverse of the sample set's cardinal number. Hence, the probability of any event is the ratio of the event's cardinal number by the sample set's cardinal number:

$$\frac{\#(\text{positive cases})}{\#(\text{possible cases})}$$

Solving a problem under the assumption of equiprobability consists in solving a counting problem: determine both numbers in the ratio.

2.2 Exercises

Exercise 2.5
From a 32-card deck, we randomly pick a 5-card hand. Determine the probabilities of getting:

1. exactly one ace;
2. at least one ace;
3. the queen of clubs;

[3] This question is the starting point of the birthday problem, cf. Wikipedia, your best friend!

4. 4 clubs, including the queen;
5. exactly one club and one queen.

Exercise 2.6
We roll two fair dice. Find the probabilities:

1. that the numbers on the dice are equal;
2. that the sum of the dice is an even number;
3. that the sum of the dice is less than or equal to 8;
4. that the highest number of the dice is less than or equal to 4.

3 Conditional probabilities and independence

3.1 Summary

When modeling a random experiment, it can happen that the probability of an event is easier to determine if we make an additional assumption. For example, it is easier to determine the probability that a person has blue eyes if we make an assumption about his parent's eyes color. Conditional probabilities enable one to define a probability ***under an assumption***. Furthermore, thanks to the law of total probability, it is possible to express the probability of an event as a sum of probabilities under assumptions, provided that these assumptions are mutually exclusive collectively exhaustive events[4].

Sometimes, the probability that an event A occurs under the assumption that another event B occurs (the «probability of A given B») is equal to the probability of A. Then, the information on whether B occurs or not does not alter the probability of A. We say that the events A and B are ***independent***. Assuming the independence of two events or two random experiments leads to a drastic simplification of the probability model.

When two events A and B are not independent, the information that A occurred alters the probability that B occurs. Conversely, the information that B occurred alters the probability that A occurs. Bayes' formula gives the relation between these two conditional probabilities. It states that

$$\mathrm{P}(A \mid B) = \frac{\mathrm{P}(B \mid A)\mathrm{P}(A)}{\mathrm{P}(B)}$$

This formula is very useful: in many situations, one of the conditional probabilities is easy to estimate, but the conditional probability that we need, the one that will help in taking a decision, is the second one.

Statistics has the purpose to use data in order to decide whether a property is true or not. Imagine a property A which is not observable: it is not possible to observe whether it is true or false. Statisticians try to collect data to infer the answer. Then they define the random event $B =$ «the data are the ones that have been observed» and use a probability model. In history, two ways of thinking have emerged: the frequentist and the Bayesian approaches. Frequentist statisticians will suppose that A is true and evaluate $\mathrm{P}(B)$: if it is not to low, there is no evidence that A is false and the property will be accepted. Bayesian statisticians will assume that A is random, will give an *a priori* value for $\mathrm{P}(A)$ and will use Bayes' formula to evaluate $\mathrm{P}(A \mid B)$.

[4]Then they are a ***partition*** of the sample set.

3.2 Exercises

Exercise 3.7

1. Assume you randomly pick a number in $[\![1, 20]\!]$. Consider the two events:

$$A = \text{"the number is even"} \qquad \text{and} \qquad B = \text{"the number is a multiple of 5"}$$

   (a) Are the events A and B incompatible?
   (b) Are the events A and B independent?

2. Answer the previous questions if we pick the number in $[\![1, 21]\!]$.

Exercise 3.8

Which statement is given by the law of total probability? Prove it[5].

Exercise 3.9

In a commercial printing company, some posters are printed with a slight defect in coloring. The proportion of such defective posters is 0.05. The director of the company set up a quality checking service whose results are the following: a perfect poster has the probability 0.96 of being accepted while a defective poster has the probability 0.98 of being rejected.

1. Interpret these data as probabilities or conditional probabilities of simple events.
2. Determine the probabilities of the following events:
   (a) there is an error from the quality checking service;
   (b) an accepted poster is defective.

Exercise 3.10

Let A and B be two dice: dice A has 4 red sides and 2 white sides while dice B has 2 red sides and 4 white sides. To start with, we toss a loaded coin, whose probability of getting tails is $1/3$, in order to determine which dice will be used: if the coin toss is tails, we use dice A, if it is heads we use dice B. Then, we roll the selected dice several times.

1. Determine the probability of having a red side at the first roll of the dice.
2. Let $n \in \mathbb{N}^*$ and suppose that the first n rolls of the dice all resulted in a red side. What is the probability that dice A has been used?
3. Suppose that the first two rolls of the dice both resulted in a red side. What is the probability of having a red side at the third roll?

Express each answer in two ways: (a) with a formula involving conditional probabilities; (b) with a numerical simplified fraction.

Exercise 3.11

In order to get ANAC points (they are required to validate the S2 semester), EPITA students have two possibilities: doing sport or getting involved in communication activities. It has been estimated that 60% out of the students get points with sport and 50% with communication activities. Furthermore, one third of the students doing sport also do communication activities.

Represent these data with a table. Determine the percentage of students who will have to repeat their S2 because they will have got no ANAC points.

[5] Most of the time, we forget this law. We just prove it when reasoning.

4 Finite random variables

4.1 Summary

The distribution of a random variable can be highlighted with localization's and dispersion's descriptors. In statistics, the localization of a data set is given by its average. A similar descriptor in probability is the ***expectation***. The ***variance*** and the ***standard deviation*** describe the dispersion of the variable around its expectation. These descriptors are related to the first two moments of the variable, those of orders 1 and 2. The set of all the moments of a variable (orders 1, 2, 3, ...) characterizes the distribution. Hence, in some cases, a complicated distribution is approximated by a simpler one, whose first moments are equal (or close) to those of the initial distribution[6]. Then the slight loss of accuracy is compensated by easier computations. In the following, we will focus on the first two moments, that is, the expectation and the variance of a random variable.

4.2 Exercises

Exercise 4.12

1. Let X be a random variable taking a finite number of values. Remind how its expectation and variance are computed. Explain these notions. What formula do you remember about the expectation and variance of finite variables?
2. We roll a fair dice. Let us define the random variables:
   - T, the number on the dice.
   - X, the double of the number on the dice, minus 6.

   (a) Determine the distribution of T. Deduce $\mathrm{E}(T)$ and $\mathrm{V}(T)$.
   (b) What are the possible values for X? Determine a relation between variables X and T.
   (c) Deduce $\mathrm{E}(X)$ and $\mathrm{V}(X)$.

Exercise 4.13

A box contains:

- four blue balls numbered 1 to 4,
- three red balls numbered 1 to 3,
- two green balls numbered 1 to 2.

You randomly pick one ball.

1. Let X be the random variable $X =$ "number of the ball".

   Find the distribution of X, its expectation and its variance.
2. Assume you play the following game: you randomly pick one ball.
   - If the ball is blue and has an even number, you win 2 euros.
   - If the ball is not blue and has an even number, you win 3 euros.
   - Otherwise, you win nothing.

   Let Y be your gain at the game.

   (a) What is the distribution of Y?

[6] For example, in suitable conditions (see exercise 5-21), a hypergeometric distribution can be replaced with a binomial. And the binomial distribution can be replaced with a Poisson distribution.

(b) What is the probability that you win some money?

Exercise 4.14
From a 32-card deck, we randomly pick a 5-card hand. Let X be the number of hearts in the hand.

1. What is $X(\Omega)$, the set of possible values for X?
2. Write the distribution of X.
3. Compute its expectation $\mathrm{E}(X)$. You may use the "Vandermonde formula": for all $(a, b, n) \in \mathbb{N}^3$ such that $n \leqslant a + b$,
$$\sum_{k=0}^{n} \binom{a}{k}\binom{b}{n-k} = \binom{a+b}{n}$$
4. ***(Bonus)*** Using a counting problem, prove the Vandermonde formula.

Exercise 4.15
In statistical analysis, a common way to preprocess the data consists in replacing the variables with their ***standardized variables***. For examples, many machine learning algorithms work better when using standardized data.

Définition 4.1. *If X admits an expectation and a variance*[7] *and if its variance is non-zero*[8]*, then the standardized variable of X is the variable*

$$\overline{X} = \frac{X - E(X)}{\sigma_X}.$$

Show that $\mathrm{E}(\overline{X}) = 0$ and $\mathrm{V}(\overline{X}) = 1$.

Exercise 4.16
1. From a 32-card deck, we randomly pick one card. Let X and Y be the random variables

$$X = \text{«number of queens»} \qquad \text{and} \qquad Y = \text{«number of hearts»}$$

   a. Find the distributions of X and Y.
   b. Represent the distribution of the 2-tuple (X, Y) with a table.
   c. Show that X and Y are independent.

2. Now, we pick two cards.

   a. Find the distributions of X and Y.
   b. Represent the distribution of the 2-tuple (X, Y) with a table.
   c. Show that X and Y are not independent.

5 Typical distributions

5.1 Summary

Some typical random experiments, such as picking cards from a deck, winning or losing at a game, appear in many modeling processes. These experiments can be described by random variables, whose distributions are the «typical distributions». This section presents some of them: the uniform, Bernoulli, binomial and hypergeometric (see exercise 4-14) distributions. .

[7]Any finite variable admits an expectation and a variance. However, some infinite variables do not.
[8]For a finite variable, this condition just means that it is a non-constant variable.

5.2 Exercises

Exercise 5.17

Let X be a uniform-distributed variable on $[\![0, n]\!]$.

1. Compute its expectation.
2. Find its variance.

N.B.: remind that, during the seminar, we have proven the properties:

$$\sum_{k=1}^{n} k = \frac{n(n+1)}{2} \quad \text{and} \quad \sum_{k=1}^{n} k^2 = \frac{n(n+1)(2n+1)}{6}$$

Exercise 5.18

1. What is the expectation of a Bernoulli-distributed variable, with parameter $p \in [0, 1]$?
2. Deduce the expectation of a binomial-distributed variable with parameters $n \in \mathbb{N}$ and $p \in [0, 1]$.
3. In the previous two questions, replace the expectation with the variance.

Exercise 5.19

An EPITA maths teacher has just collected the worksheets of his 40 students' midterm exam. Now, he must read them and give a mark. Being lazy, he makes use of the best method for this work: rolling a 20-sided dice, each side being labeled 1 to 20.

1. For each student $i \in [\![1, 40]\!]$, the teacher rolls the dice. The result M_i is the mark that the student gets. What is the distribution of M_i? Its expectation? What is the probability that student i gets 10 or more?
2. Let X be the number of students who get 10 or more. What is the distribution of X?

Exercise 5.20

The program for a math exam consists of three chapters. However, the teacher feels lazy, he wants to enjoy the last sunny days before the fall. Hence, he decides to write only one exercise, covering one random chapter among the three.

One student is as lazy as the teacher; he also wants to enjoy the last sunny days. Therefore, he chooses to review only two chapters among the three.

If the chapter that is covered by the exam is, fortunately, one of the two that the student has reviewed, then his probability of validating the semester is 80%. Otherwise, this probability is only 20%.

1. Represent this problem graphically.
2. What is the probability that the student validates the semester?
3. Two weeks later, the marks are known and the student has validated the semester. What is the probability that he had reviewed the chapter covered by the exam exercise?
4. Suppose now that the 40 students from the class choose this method for reviewing their exam. Let X be the number of students validating the semester. What is the distribution of X? Determine its expectation.

Exercise 5.21

1. In the Paris area, at each birth, the probability of having a boy is 46%. For a family of 5 children, determine the probability of having 2 boys and 3 girls.

2. Consider now a sample of 100 children, 46 of them being boys, 54 being girls. We randomly select 5 children. What is the probability of getting 2 boys and 3 girls?

3. Explain why these two questions lead to different models and results.