Probability Calculus

Josemari Sarasola

Statistics for Business

Gizapedia

Combinatorics

What is combinatorics?

- Before learning to calculate probabilities, we must learn to count.
- Combinatorics is the collection of mathematical methods to count in complex problems.
- Question: how many subsets of 2 elements can we create from a set of 4 elements (a,b,c,d)?

Ways to count

In order to give an answer to the previous question, we must specify which subsets of 2 elements are *really* different:

- If we account the order and don't accept repetitions, we have 12 pairs: *ab-ac-ad-bc-bd-cd-ba-ca-da-cb-db-dc*. These are *variations without repetition*.
- If we account the order and do accept repetitions, we have 16 pairs: add *aa-bb-cc-dd* to the previous pairs. These are *variations with repetition*.
- If we don't consider the order and don't accept repetitions, we have 6 pairs: *ab-ac-ad-bc-bd-cd*. These are *combinations*.
- If we don't consider the order and accept repetitions, we have 10 pairs: add *aa-bb-cc-dd* to the previous pairs. These are *multicombinations*.

Combinatorial formulas

In the general problem, we have to choose a subset of k elements from a main set of size n. These are the general formulas to count in the ways we have previously given:

- Variations without repetition (also called k-permutations of n): $V_n^k = \dfrac{n!}{(n-k)!}$
- Variations with repetition : $VR_n^k = n^k$
- Combinations: $C_n^k = \dbinom{n}{k} = \dfrac{n!}{k!(n-k)!}$
- Multicombinations: $M_n^k = \dbinom{n+k-1}{k}$

Exercise: Apply these formulas in order to calculate the results in the previous slide.

Combinations: properties

- $\binom{n}{0} = 1$
- $\binom{n}{1} = n$
- $\binom{n}{n-k} = \binom{n}{k}$
- Shortcut: in order to calculate $\binom{n}{k}$, multiply n in a "decreasing way" just the number of times shown by number k, and divide it by $k!$. E.g.: $\binom{7}{3} = \frac{7 \times 6 \times 5}{3!} = 35.$

Another way to count: permutations

- In other problems we want just to arrange n elements in different orders. Each of these ordered sequences is called a permutation.
- The number of permutations of n elements is $P_n = n!$
- E.g.: a-b-c elements are ordered in 3!=6 ways: abc, acb, bac, bca, cab, cba.

Permutations with repetition

- But the previous formula doesn't work when some of the elements are repeated (e.g, with a-a-b elements).
- In those cases, we must apply the formula for permutations with repetition (α, β: number of repetitions):

$$PR_n^{\alpha,\beta,\ldots} = \frac{n!}{\alpha!\beta!\ldots}$$

- E.g.: a-a-b elements are ordered in 3!/2!1!=3 ways: aab, aba, baa.

Probability

Probability: definition and interpretation

- Probability: measure of uncertainty of a random event.
- E.g.: the probability of raining in a day in August in Donostia is 0.2 (or 20%)
- Interpretation: in the long term, around in 20% of the days in August it will rain.

Probability

Basic methods to calculate probability

- Laplace's rule
- Frequential interpretation
- Subjective probability

Laplace's rule

- $P(A) = \dfrac{number\ of\ outcomes\ resulting\ into\ A}{total\ number\ of\ outcomes}$
- Example: Probability of getting odd after throwing a dice:

$$P[odd] = \frac{\{1,3,5\}}{\{1,2,3,4,5,6\}} = 0.5$$

- BUT, all possible outcomes must be EQUIPROBABLE (with the same chances to happen) to apply this rule.
- Example: Can we apply the rule for the probability of raining tomorrow?
- When we apply Laplace's rule, we often must use combinatorial formulas, in order to count the number of possible outcomes.

Frequential interpretation

- Example: how to calculate the probability for an abstract new student to pass the statistics exam?
- According to Laplace's rule, it will 0.5, as there are two outcomes: *passsing* and *not passing*.
- BUT, we cannot take those two outcomes as the were EQUIPROBABLE. So, what can we do?
- Compile data from the past or make some experiments.
- Example: We give an exam to 200 new students, and 160 (frequency for passing) of them passed. So:

$$P[passing] = \frac{160}{200} = 0.8$$

Frequential interpretation: remarks

- Experiments and data must be homogeneous, that is, always with the same conditions (don't pass the same exam in different days!)
- It can be expensive and difficult to undertake (experiments in nuclear plants, ummm!)
- The result is an estimation, but a big number of experiments will give us a small error almost surely.

Subjective probability

- Example: What is the probability of you (yes, you) passing the statistics exam?
- We cannot carry out experiments with you, because as you make exams you are better in statistics. So, it's impossible to repeat the same conditions.
- Solution: take into account all personal factors about learning, understanding and studying (Are you clever? Did you understand the course? Do you regularly study?) and according to the results, make a a subjective estimation of the probability.
- Another examples: probability of inflation rising next year, probability of sales increasing next year, probability of Euskal Herria being independent from Spain in the next ten years.

Subjective probability

Generally, subjective probability is the level of belief that a rational individual has about the ocurrence of a given event. The subjektive probability can be given in two ways:

- **weak subjectivism**: when objective factors (causality, experience) are taken into account in roder to quantify the probability (we also this interpretation, *epistemic probability*).
- **strong subjectivism**: when there is no objective factor to give the probability, and the probability relies absolutely on the beliefs of the individual.

Probability: event algebra

Event algebra

- We have learned how to calculate probability of simple *events*. But how to calculate the probability of combined events? And, how to combine events?
- So, now we are going to learn *event algebra*.
- First concept: *sample space* or *universe* , named Ω, the set of possible outcomes of a random process.
- Example, afther throwing a dice: $\Omega = \{1, 2, 3, 4, 6\}$ or $\Omega = \{odd, even\}$.
- We usually depict the sample space by means of a Venn diagram:

$$\boxed{\begin{array}{cccccc} & 2 & & 4 & & 6 \\ 1 & & 3 & & 5 & \end{array}}_{\,\Omega}$$

Complementary events

- The complementary event of any A event is the event that happens when A doesn't occur. We write it $\overline{A}$
- Example: about gender, man and woman.
- Depiction:

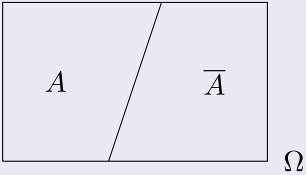


Figure: A eta $\overline{A}$ are complementary events.

- Basic rule for complementary events: $P[\overline{A}] = 1 - P[A]$

Probability: event algebra

Mutually exclusive events

- Two events are mutually exclusive if they can't occur at the same time.
- Depiction:

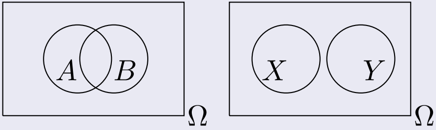


Figure: A eta B (on the left) are not mutually exclusive. X eta Y (on the right) are mutually exclusive.

- Example: *Man* and *using tampons* are mutually exclusive, bu man and vegetarian are not.
- Exercise: Are complementary events mutually exclusive? And inversely?

Probability: event algebra

Intersection of events

- The intersection of two events A and B, denoted by $A \cap B$, is the event that happens when both A and B happen at the same time.
- Depiction:

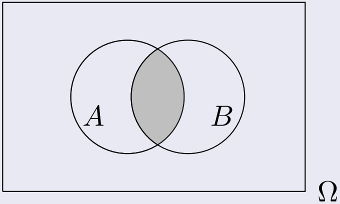


Figure: $A \cap B$ (gray): The intersection of A and B events

- Exercise: How is the intersection of two complementary events? And what is its probability?

Union of events

- The union of two events A and B, denoted by $A \cup B$, is the event that happens when at least one of those events happen.
- Depiction:

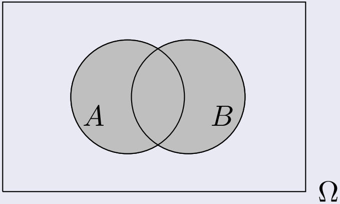


Figure: $A \cup B$ (gray): The union of A and B events

- When $A_1, A_2, \ldots, A_n$ are mutually exclusive, here is how we calculate the probability of the their union:

$$P(A_1 \cup A_2 \cup \ldots \cup A_n) = P(A_1) + P(A_2) + \ldots + P(A_n)$$

Substraction of events

- The substraction of B event from A event , denoted by $A - B$, is the event that happens when A happens but B not.
- Depiction:

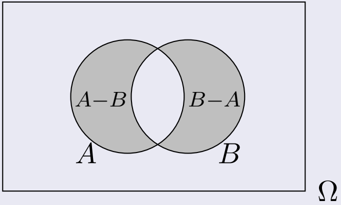


Figure: $A - B$: "A happens but B not" and $B - A$: "B happens but A not".

Probability: event algebra

Events as subsets of other events

- A is a *subset* of B event , denoted by $A \subset B$, if whenever A happens, B also happens.
- Depiction:

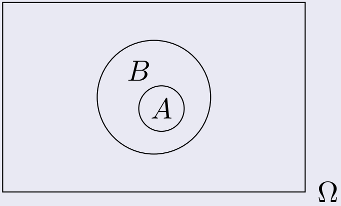


Figure: $A \subset B$: A is a subset of B.

- If A is subset of B, what is $A \cap B$? And $A \cup B$?

Probability: event algebra

De Morgan's laws

- $\overline{A \cup B} = \overline{A} \cap \overline{B}$
- $\overline{A \cap B} = \overline{A} \cup \overline{B}$

Probability: event algebra

Inclusion-exclusion principle

We apply the inclusion-exclusion principle to calculate the probability of the union of events

- For A and B events: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- For, A, B and C events:

$$\begin{aligned} P(A \cup B \cup C) =& P(A) + P(B) + P(C) \\ &- P(A \cap B) - P(A \cap C) - P(B \cap C) \\ &+ P(A \cap B \cap C) \end{aligned}$$

Inclusion-exclusion principle

For $A_1,\ A_2,\ A_3,\ \ldots,\ A_n$ events:

$$
\begin{gathered}
P(A_1 \cup A_2 \cup A_3 \cup \ldots \cup A_n) = \\
P(A_1) + P(A_2) + P(A_3) + \ldots + P(A_n) \\
-P(A_1 \cap A_2) - P(A_1 \cap A_3) - \ldots - P(A_{n-1} \cap A_n) \\
+P(A_1 \cap A_2 \cap A_3) + \ldots + P(A_{n-2} \cap A_{n-1} \cap A_n) \\
-P(A_1 \cap A_2 \cap A_3 \cap A_4) - \ldots \\
+ \ldots
\end{gathered}
$$

So we must add all simple events, substract all double events, add all triple events, substract all quadruple events, and so on, till reaching the whole set of events.

Probability: event algebra

Inclusion-exclusion principle

When all events are mutually esclusive, all their possibl intersections are empty, so the inclusion-exclusion principle gives the rule we already know:

$$P(A_1 \cup A_2 \cup \ldots \cup A_n) = P(A_1) + P(A_2) + \ldots + P(A_n)$$

So whenever events are mutually exclusive, we calculate the probability of their union by just adding their simple probabilities.

Conditional probability

We denote $P(B/A)$ the conditional probability of B given A, that is, the probability of B provided (under the condition) that A has occurred:

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

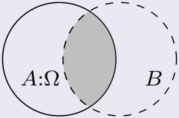


Figure: If A has occurred, sample space narrows to A event.
In that new universe, the probability of B, that is $P(B/A)$,
is the domain that B event takes into the domain of A,
that is $P(A \cap B)$, but restricted to $P(A)$.

Probability: product of probabilities

Chain rule

- Also named general product rule, it gives the probability of the intersection of two or more events.
- For two events: $P(A \cap B) = P(A) \times P(B/A)$
- For n events:
  $P(A_1 \cap A_2 \cap A_3 \cap \ldots \cap A_n) =$
  $P(A_1) \times P(A_2/A_1) \times P(A_3/A_1 \cap A_2) \times \ldots \times$
  $P(A_n/A_1 \cap A_2 \cap A_3 \cap \ldots \cap A_{n-1})$
- When events are given in a chronological order, as usual, the application of the rule is much easier when we follow that order by multiplying the probabilities.

Probability: product of probabilities

Chain rule: example

We have made a survey among Basque young people aged 18-30:

	Are you vegan?		Total
Sex	Yes	No	
Male	25	20	45
Female	40	15	55
Total	65	35	100

1. If you are a woman, what is the probability of being vegan?
   Answer:
2. We select randomly a Basque young, what is the probability of being a vegan male?
   Answer:

Probability: product of probabilities

Dependence and independence

- We say A and B are *independent* events, when knowing that one of them has happened (or not) doesn't give any information about the probability of the other one:

$$P(A \cap B) = P(A) \times P(B/A) = P(A) \times P(B)$$

- On the other side, if the occurence of one of the events gives information about the other(s), those events are *dependent*.

Probability: product of probabilities

Dependence: sampling without devolution

- Example: In an urn, we have 6 faultless and 4 faulty items. We select randomly two items without devolution. What is the probability of both of them being faultless?

$$P[1o \cap 2o] = P[1o] \times P[2o/1o] = \frac{6}{10} \times \frac{5}{9}$$

So, sampling without devolution implies dependence, as number of items changes as we extract the elements in the population.

Probability: product of probabilities

Dependence: sampling with devolution

- Example: In an urn, we have 6 faultless and 4 faulty items. We select randomly two items with devolution. What is the probability of both of them being faultless?

$$P[1o \cap 2o] = P[1o] \times P[2o] = \frac{6}{10} \times \frac{6}{10}$$

We can see probability doesn't change as we select items. No matter which are the selected elements, probability remains always constant. So, sampling with devolution implies independence, as the population remains always unchanged.

Probability: product of probabilities

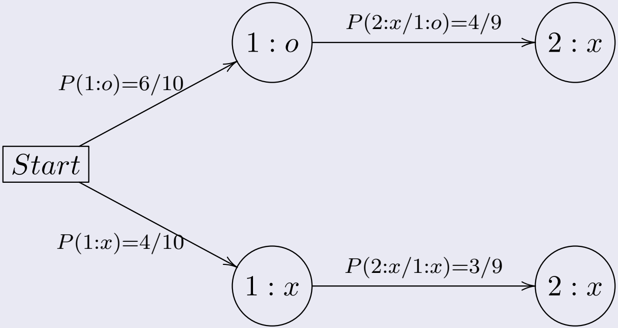
Probability trees

- Example: In an urn, we have 6 faultless and 4 faulty items. We select randomly two items without devolution. What is the probability of the second one being faulty?
- It depends:
  - If the first item is faulty: 3/9.
  - If the first item is faultless, 4/9.
- Provided we know nothing about the first item, the answer will be between the two previous values, weighted by the probabilities of the first item being faulty and faultless:
  $P(2x) = P[(1o \cap 2x) \cup (1x \cap 2x)] =$
  $P(1o) \times P(2x/1o) + P(1x) \times P(2x/1x) =$
  $\frac{6}{10} \times \frac{4}{9} + \frac{4}{10} \times \frac{3}{9}$

Probability: product of probabilities

Probability trees

Graphically it's easier:



Probability: product of probabilities

Probability trees

We have just to follow the path in order to reach the event we want,

- multiplying probabilities into the path;
- adding probabilities of different paths.

$$\frac{6}{10} \times \frac{4}{9} + \frac{4}{10} \times \frac{3}{9}$$

Some basic rules:

- We have to calculate each probability assumed that previous events have happened.
- The probabilities following a node or switch must add to 1, except at the last one, that is, except when we reach to the desired probability.

Probability: product of probabilities

Probability trees

Probability trees are useful in these situations:

- When the solution to a probability problem is 'it depends on'. In that case we switch to the different possibilities.
- When we have a chronology: a sequence of events over time. In that case, we follow the paths in the chronologic order.

Probability trees are the rule-of-thumb version of the law of total probability:

$$P(A) = \sum_n P(A/B_n)P(B_n)$$

Bayes' theorem: example

In an urn we have 8 faultless and 3 faulty items. We sent a piece randomly to a customer. After that, we inspected an item and we saw it was faulty. What is the probability of the first one being faulty?

B: 2nd faulty

A_i	$P(A_i)$	$P(B/A_i)$	$P(A_i) \times P(B/A_i)$	$P(A_i/B)$
1st faulty	$3/11{=}0.27$	$2/10{=}0.2$	0.054	0.199
1st faultless	$8/11{=}0.73$	$3/10{=}0.3$	0.219	0.801
	1		0.273	1

Probability: Bayes' theorem

Bayes' theorem: some clues

- The goal of the Bayes' theorem is to adjust the probability of an event according to an information we have got.
- The probabilities before knowing the information are $P(A_i)$ and we call them *a priori* (previous) probabilities. The adjusted probabilities according to the known facts ($P(A_i/B)$) are called *a posteriori* (following) probabilities, denoting by B the information we get and use to adjust a priori probabilities.
- The verosimilities are the probabilities of the information we get known that each of the A_i events has occurred: $P(B/A_i)$
- Most often data in a Bayes' problem are the a priori probabilities and verosimilities (first three columns) and the goal is to calculate the a posteriori probabilites.
- The sum of both a priori and a posteriori probabilities is always 1 (they represent the probabilites of everything that may happen).

Probability: Bayes' theorem

Bayes' theorem: returning to the example

- We want to know the probability of the 1st piece being faulty, so we set $P(A)$ as "faulty" and "faultless" (everything that may happen).
- We set verosimilities: if the 1st item is faulty (faultless) what is the probability of the 2nd one being faulty (the information we get).
- The following steps are all "mechanical", leading us to the a posteriori probabilites, "1st faulty" and "1st faultless" but taking b into acconut this time.
- As the 2nd piece was faulty, intuitively it's clear that the probability of the 1st piece being also faulty must be lower $(0.27 \to 0.199)$.

Example: Planet Nine

- 2016: astronomers Mike Brown and Konstantin Batygin inferred the existence of a large planet beyond Neptune: the Planet Nine.
- Set a hypothesis that can be assumed (H_0): there's no planet beyond Neptune (let's be cautious: we have not detected anything!).
- Evidence: Brown and Batygin found orbits of some TNOs (Trans Neptunian Objects) were clustered.
- Statistical testing: probability of that clustering is very small provided H_0. By contrast that clustering is likely if Planet Nine exists.
- Conclusion: we may accept the hypothesis of the Planet Nine. But it remains always a hypothesis!

Probability: introduction to statistical testing

Statistical testing: steps

- Goal: take a decision about if a hypothesis is true or not, according to evidence.
- First step: set a null hypothesis (the statement we accept as a matter of principle, with caution and without any other evidence), denoted by H_0 .
- Calculate the probability of the evidence, provided that H_0 is true.
- If the probability of the evidence (called p-value) is very low, we conclude that it's strange that the evidence happened provided that H_0 is true, so we reject H_0. Otherwise (when the p-value is not low enough) we must think the evidence is a normal fact under H_0, so we must accept H_0.
- How do we know p-value is low or not low enough? We set at the beginning (in order to be neutral) the significance level, denoted by α. It's a low enough probability; usually 0.01, 0.05 or 0.10. So, if $p - value < \alpha$, we reject H_0; otherwise we accept it.