# 2

# Concerning Certainty and Uncertainty

## 2.1 Certainty and Uncertainty

2.1.1. In almost all circumstances, and at all times, we all find ourselves in a state of uncertainty.

Uncertainty in every sense.

Uncertainty about actual situations, past and present (this might stem from either a lack of knowledge and information, or from the incompleteness or unreliability of the information at our disposal; it might also stem from a failure of memory, either ours or someone else's, to provide a convincing recollection of these situations).

Uncertainty in foresight: this would not be eliminated or diminished even if we accepted, in its most absolute form, the principle of determinism; in any case, this is no longer in fashion. In fact, the above-mentioned insufficient knowledge of the initial situation and of the presumed laws would remain. Even if we assume that such insufficiency is eliminated, the practical impossibility of calculating without the aid of Laplace's demon would remain.

Uncertainty in the face of decisions: more than ever in this case, compounded by the fact that decisions have to be based on knowledge of the actual situation, which is itself uncertain, to be guided by the prevision of uncontrollable events, and to aim for certain desirable effects of the decisions themselves, these also being uncertain.

Even in the field of tautology (i.e. of what is true or false by mere definition, independently of any contingent circumstances), we always find ourselves in a state of uncertainty. In fact, even a single verification of a tautological truth (for instance, of what is the seventh, or billionth, decimal place of $\pi$, or of what are the necessary or sufficient conditions for a given assertion) can turn out to be, at a given moment, to a greater or lesser extent accessible or affected with error, or to be just a doubtful memory.

2.1.2. It would therefore seem natural that the customary modes of thinking, reasoning and deciding should hinge explicitly and systematically on the factor *uncertainty* as the conceptually pre-eminent and determinative element. The opposite happens, however: there is no lack of expressions referring to uncertainty (like 'I think', 'I suppose', 'perhaps', 'with difficulty', 'I believe', 'I consider it as probable', 'I think of it as likely', 'I would bet', 'I'm almost certain', etc.), but it seems that these expressions, by and large, are no more than verbal padding. The solid, serious, effective and essential part of arguments, on the other hand, would be the nucleus that can be brought within the

*Theory of Probability: A Critical Introductory Treatment*, First Edition. Bruno de Finetti.

language of certainty – of what is certainly *true*, or certainly *false*. It is in this ambit that our faculty of reasoning is exercised, habitually, intuitively and often unconsciously.

In reasoning, as in every other activity, it is, of course, easy to fall into error. In order to reduce this risk, at least to some extent, it is useful to support intuition with suitable superstructures: in this case, the superstructure is *logic* (or, to be precise, the *logic of certainty*).

Whether it is a question of traditional verbalistic logic, or of mathematical logic, or of mathematics as a whole, the only difference in this respect is in the degree of extension, effectiveness and elegance. In fact, it is, in any case, a question of ascertaining the *coherence*, the *compatibility*, of stating, believing, or imagining as hypotheses some set of 'truths'. To put it in a different way: thinking of a subset of these 'truths' as *given* (knowing, for instance, that certain facts are true, certain quantities have given values, or values in between given limits, certain shapes, bodies or graphs of given phenomena enjoy given properties, and so on), we will be able to ascertain which conclusions, among those of interest, will turn out to be – on the basis of the data – either *certain* (certainly true), or *impossible* (certainly false), or else *possible*. The qualification 'possible' – which is an intermediate, generic and purely negative qualification – is applied to everything that does not fall into the two extreme limit cases: that is to say, it expresses one's ignorance in the sense that, on the basis of what we know, the given assertion could turn out to be either true or false.

2.1.3. This definition of 'possible' itself reveals an excessive and illusory confidence in 'certainty': in fact, it assumes that logic is always sufficient to separate clearly that which is determined (either true or false), on the basis of given knowledge, from that which is not. On the contrary (even apart from the possibility of deductions which are wrong, or whose correctness is in doubt), to the sphere of the *logically possible* (as defined above) one will always add, in practice, a fringe (not easily definable) of the *personally possible*; that is that which must be considered so, since it has not been established either that it is a consequence of one's knowledge or that it is in conflict with it.

We have already said, in fact, that logic can *reduce the risk of error*, but cannot eliminate it, and that tautological truths are not necessarily accessible. However, in order not to complicate things more than is required to guard against logical slips, we will always consider the case in which 'possible' can be interpreted as *logically possible*.[1]

## 2.2 Concerning Probability

2.2.1. The distinction between that which, at a certain moment, we are ignorant of, and that which, on the other hand, turns out to be certain or impossible, allows us to think about the range of *possibility*; that is, the range over which our uncertainty extends. However, this is not sufficient as an instrument and guide for orientation, decision or action: to this end –and this is what we are interested in – it will be necessary to base oneself on a further concept; the concept of *probability*.

1 Possibly by *eliminating* some knowledge. For instance, in the case of $\pi$ it seems reasonable (for the problem under consideration) to imagine that one ignores the properties that permit the calculation of $\pi$, and to consider it as an 'experimental constant' whose decimal representation could only be known if somebody had determined it and published the result. I believe that for a mathematician, too, it would be reasonable to think that everything proceeds as if he were in such a state of ignorance.

In this chapter we do not wish to talk about probabilities, however; they will be introduced in Chapter 3. This deferment is undoubtedly awkward: obviously, the awkwardness consists in introducing preliminary notions without, at the same time, exhibiting their use. Didactically this is a bad mistake – one runs the risk of making boring and dull that which otherwise would appear clear and interesting. However, when it is important to emphasize an essential distinction, which otherwise would remain unnoticed and confused, a rigid separation is necessary – even if it seems to be artificial and pedantic. This is precisely the case here.

2.2.2. The study of the range of possibility, to which we shall here limit ourselves, involves learning how to know and recognize all that can be said concerning uncertainty, while remaining in the domain of the logic of certainty; that is, in the domain of what is *objective*. Probability will be a further notion not belonging to that domain and, therefore, a *subjective* notion. Unfortunately, these two adjectives anticipate a question about which there could be controversial opinions – their use here is not intended to prejudice the conclusion, however. For the time being, what matters is to make clear a distinction that is methodologically fundamental: afterwards, one can discuss the interpretation of the meaning of the two fields it delineates, the choice of nomenclature, and the points of view corresponding to them. It is precisely in order to be able to discuss them lucidly *afterwards* that it is necessary to avoid an immediate discussion of possibility and probability together; the confusion so formed would be difficult to resolve.

Both the distinction and the connection between the two fields are easily clarified: the logic of certainty furnishes us with the range of possibility (and the 'possible' has no gradations); probability is an additional notion that one applies within the range of possibility, thus giving rise to gradations ('more or less probable') that are meaningless in the logic of certainty.

2.2.3. Since it is certain that everyone knows enough about probability to be able to interpret these explanations in a less vague fashion, we can say that 'probability is something that can be distributed over the field of possibility'. Using a visual image, which at a later stage might be taken as an actual representation, we could say that the logic of certainty reveals to us a space in which the range of possibilities is seen in outline, whereas the logic of the probable will fill in this blank outline by considering a mass distributed upon it.

There is no harm in anticipating the developments that the treatment will undergo from the next chapter onwards, provided that, from the fact that they are not talked about here, one understands that they do not belong in the domain that we now consider it important to present as well-delimited and distinct.

## 2.3 The Range of Possibility

2.3.1. *Prologue.* Let us introduce right away the use of 'You', following Good (Savage uses 'Thou'). The characterization of what is *possible* depends on the state of information. The state of information will be that (at a given moment) of a real individual, or it might even be useful to think of a fictitious individual (as an aid to fixing ideas). This individual, real or fictitious, in whose state of information – and, complementarily, of *uncertainty* – we are interested, we will denote by 'You'. We do so in order that You, the reader, can better identify yourself with the rôle of this character. This character – or, better, You – will play a much more important rôle after this chapter, when probabilities will enter the scene. For the moment, You are in the audience, because You have to limit yourself to passively recording what You know for certain, or what You do not know.[2] All the same, it will be useful for You to at least get used to putting yourself in this character's place, since, even if it is not yet time to speak our lines, we are about to walk onto the stage – that is to enter into the range of possible alternatives.

With regard to any situation or problem that You have to consider, there will always exist an enormous number of conceivable alternatives. Your information and knowledge will, in general, permit You to exclude some of them as impossible: that is, they will permit You – and this has been said to be the function of science – a 'limitation of expectations.' All the others will remain *possible* for You; neither certainly true, nor certainly false. It will not happen that only one of them will be isolated as *certain*, except in special cases, or unless a rather crude analysis of the situation is given. Obviously, it is always sufficient to take all the possible alternatives and present them as a whole in order to obtain a single alternative which is 'certain.'

The choice of which of the more or less sophisticated, detailed, particularized forms we need, or consider appropriate, in order to distinguish or subdivide such alternatives, according to the problems and the degree of refinement we require in considering them, depends on us, on our judgment. Also, we have available several possible languages in which we can express ourselves in this connection. It is convenient to introduce them straight away, and altogether, in order to show, at the same time, on the one hand their essential equivalence, and, on the other, the differences between them which render their use more or less appropriate in different cases.

2.3.2. *Random events and entities*. Everything can be expressed in terms of *events* (which is the simplest notion); everything can be expressed in terms of *random entities* (which is the most generic and general notion); and so on. One or other of these notions is sufficient as a starting point to obtain all of them. However, it is instructive to concentrate attention on four notions which immediately allow us to frame within the general scheme the most significant types of problems, important from both the conceptual and practical points of view.

We will consider:

*random events*,
*random quantities*,
*random functions*,
*random entities*.

Let us make clear the meaning that we give to 'random': it is simply that of 'not known' (for You), and consequently 'uncertain' (for You), but *well-determined* in itself. Not even the circumstance of 'not known' is to be taken as obligatory; in the same way we could number constants among functions, though we will not call a constant a 'function' if there is no good reason. To say that it is *well-determined* means that it is *unequivocally individuated*. To explain this in a more concrete fashion; it must be specified in such a way that a possible bet (or insurance) based upon it can be decided without question.

2 You would have a more personal and autonomous rôle if we took into account the faculty, which You certainly possess, of considering as 'possible' that which You could show to be impossible, but which demands too much deductive effort. However, we have stated, in Section 2.1.3, that, for the sake of simplicity, we omit consideration of such hypotheses.

2.3.3. First, let us consider *random quantities*: this is an intermediate case from which we can pass more easily to the others, particularizing or generalizing as the case may be. We will denote a number, considered as a random quantity, by a capital letter; for example $X$ or $Y$, and so on. It might be an integer, a real number, or even a complex number; but the latter case should be specified explicitly. The true value is unique, but if You call it random (in a nonredundant usage) this means that You do not know the true value. Therefore, You are in doubt between at least two values (possible for You), and, in general, more than two – a finite or infinite number (for instance, all the values of an interval, or all the real numbers). We will denote by $I(X)$ the set of possible values of $X$, and we will write, in abbreviated form, $\inf X$ and $\sup X$ for $\inf I(X)$ and $\sup I(X)$. It is particularly important to distinguish the cases of random quantities which are *bounded* (*from above and below*), that is $\inf X$ and $\sup X$ finite, and those which are only *bounded from above*, or only *bounded from below*, or *unbounded*, that is $\inf X = -\infty$, or $\sup X = +\infty$, or both.

To exemplify what we mean by *well determined* in the case of random quantities, let us put $X$ = the year of death of Cesare Battisti.[3] The true value is $X = 1916$. While he was alive this value was not known to anyone and all years from that time on were possible values (for everybody). After the event, it is only random for those who are ignorant of it: for instance, for those who know only that it happened during Italy's participation in the World War I, the possible values are the four years 1915, 1916, 1917 and 1918.

Every function of a random quantity, $Y = f(X)$ or of two (or more), $Z = f(X, Y)$, and so on, is a random quantity (possibly 'degenerate', i.e. certain, if, for instance, $f(X)$ has the same value for all possible values of $X$).

2.3.4. An *event* (or *proposition*) admits only two values: TRUE and FALSE. In place of these two terms it is convenient to put the two values 1 and 0 (1 = TRUE, 0 = FALSE); in this way we simply reduce to a special case of the preceding, with an obvious, expressive meaning. Thus, when we wish to interpret the convention in this way, the event is identified with a gain of 1 if the event occurs and with a gain of 0 if the event does not occur. Moreover, with this convention the logical calculus of the events is simplified.

We continue to denote events with capital letters; in the main, $E, H, A, B, \ldots$. It is clear, for instance, that $1 - E$ is the *negation of* $E$, which is false if $E$ is true, and vice versa (value 0 if $E = 1$, and conversely): it is also clear that $AB$ is the logical product of $A$ and $B$, that is true if both $A$ and $B$ are true, and so on (this is merely an example, the topic will be developed later, in Section 2.5).

An event corresponds to a question which admits only two answers; YES or NO (YES = 1, NO = 0). It is clear that with a certain number of questions of this type we can obtain an answer to a question that involves any number of alternative answers. Given a *partition* into $s$ alternatives (one, and only one, of which is true), we can consider, for instance, the $s$ events (exclusive and exhaustive) which correspond to them. But even less is sufficient: with $n$ events we can imagine $2^n$ dispositions of YES–NO answers; we therefore have a partition into $s = 2^n$ alternatives if all these answers are possible, or into a smaller number, $s < 2^n$, if some of them are impossible (see Section 2.7 for further details).

Abandoning the restriction to a 'finite number', it is clear that by means of events we can study every case, even those involving an infinite number of possibilities.

3 Cesare Battisti was deputy for Trento at the Vienna Parliament; he volunteered for the Italian army, was then taken prisoner and hanged by the Austrians in 1916.
(Trento, where the author once lived, is an Italian city which was, in Battisti's time, a part of Austria.)

2.3.5. By talking about *random entities* in general, we have a means of expressing in a synthetic form the situation presented by any problem whatever. It is a question of referring oneself at all times to the same perspective, the one already implicitly introduced in the case of a random quantity, and which we now wish to make more precise and then to extend.

In the case of a random quantity, $X$, we can visualize the situation by considering as the '*space of alternatives*', $\mathscr{S}$, a line, the $x$-axis,[4] and on it the set, $\mathscr{Q}$, of the only values (points) *possible* (for You). In this way we consider *en masse*, implicitly, all the events concerning $X$ (that it belongs to a half-line, $X \leqslant x$, or to an interval, $x' \leqslant X \leqslant x''$, or to any arbitrary set, $X \in I$).[5]

But now it is obvious that the same representation holds in all cases (in a more intuitive sense, of course, in three, or fewer, dimensions). If we consider two random quantities, $X$ and $Y$, we can think of the Cartesian plane, with coordinates $x$ and $y$, as the space $\mathscr{S}$ in which we have a set $\mathscr{Q}$ of *points* (pairs of values for $X$ and $Y$) *possible* (for You) for a random *point* $(X, Y)$. Every event (proposition, statement) concerning $X$ and $Y$ corresponds to a set $I$ of $\mathscr{S}$: of course, only the intersection with $\mathscr{Q}$ is required, but it is simpler (and innocuous) to think of all sets $I$. The same could be said in the case of three random quantities $X$, $Y$, $Z$ (in this case $\mathscr{S}$ is ordinary space), or for more than three.

Independently of the coordinate system, we could, in this geometric representation, formulate a problem straightaway. It might concern a *random point* on a plane (e.g. that point which would be hit in firing at a target), or in ordinary space (e.g. the position, at a given instant, of a satellite with which we have lost contact). We find an appropriate representation for the situation of a particle (position and velocity) by using six-dimensional space: the space of dimension $6n$ serves as 'phase space' for the case of $n$ particles.

Independently of the geometrical meaning, or any meaning that suggests (in a natural way) a geometrical representation, we can always imagine, for any *random entity*, an abstract space $\mathscr{S}$ consisting of all possible alternatives (or, if convenient, a larger space of which these form a subset $\mathscr{S}$). We could consider, for example, *random vectors*, *random matrices* or *random functions*, and, thus far, the linear structure of the space continues to present itself as natural. But we could also consider *random sets*: for example, *random curves* (the path of a fly, or an aeroplane), random sets on surfaces (that part of the earth's surface in shadow at a given instant, or on which rain fell in the last 24 hours); or we could think of random entities inadequate to give any structure to the space.

We can, therefore, accept this representation as the general one, despite some reservations which will follow shortly (the latter are intended not as arguments against the representation, or for its rejection, but rather in favour of its acceptance 'with a pinch of salt').

---

4 We always denote by $x$($y$, etc.) the axis on which $X$($Y$, etc.) is represented.

5 We omit here critical questions relating to the possibility of giving, or not giving, a meaning to statements of an extremely delicate or sophisticated nature (or at least to the possibility of taking them into consideration). For example, the distinction between $<$ and $\leqslant$, the case of I 'nonmeasurable' in some sense or other, etc. It will be necessary to say something in Chapter 6; discussion of a critical character will be developed only in the Appendix, apart from brief anticipatory remarks here and there.

2.3.6. There is no need to deal with *random functions* separately, by virtue of the particular position they hold with respect to the preceding considerations (just as events and arbitrary entities have extreme positions, and random quantities an intermediate, but instrumentally fundamental, position). It is useful, however, to mention them explicitly for a moment. Firstly, in order to point out an example of applications which become more and more important from now on, and are largely new with respect to the range of problems traditionally recognized. Secondly, because we can allude, in a simple and intuitive way, to certain critical observations of the kind that will be reserved, in general, for the Appendix.

A *random function* is a function whose behaviour is unknown to You: we will denote it by $Y(t)$, assuming for convenience of intuition that the variable $t$ is time.[6] If the function is known up to certain parameters, for instance $Y(t) = A\cos(Bt + C)$ with $A$, $B$, $C$ random (i.e. unknown to You), the whole thing is trivial and reduces to the space of parameters. The case which, in general, we have in mind when we speak of a random function – or a *random process*, if we wish to place more emphasis on the phenomenon than on the mathematical translation – is that in which (to use the suggestive, if somewhat vague, phrase of Paul Lévy) the uncertainty exists at every instant (or, in his original expression, 'chance operates instant by instant').

This might mean, for example, that knowing the values of $Y(t)$ at any number of instants, $t = t_1, t_2, \ldots, t_n$, however large the (finite) $n$, the value at a different instant $t$ will still, in general, be uncertain. Sometimes, either for simplicity or in order to be 'realistic', we imagine that it makes sense to measure $Y$ at a finite (although unrestrictedly large) number of instants, without disposing of other sources of knowledge.[7] In such cases, the space $\mathscr{S}$ can be thought of as that in which every function is a 'point', but in which the possibility of distinguishing whether or not a function belongs to a set is only possible for those sets defined by a finite number of coordinates: the latter, being observable, are actually events. The simplest form of these events occurs when we ask whether or not the values at given instants fall inside fixed intervals $a_h \leqslant Y(t_h) \leqslant b_h$, $h = 1, 2, \ldots, n$. To give a visual interpretation, we ask whether or not the graph passes through a sequence of $n$ 'doors', like a *slalom*.

## 2.4 Critical Observations Concerning the 'Space of Alternatives'

2.4.1. Having reference to the 'space of alternatives' undoubtedly provides a useful overall visualization of problems. Nevertheless, the systematic and, in a certain sense, indiscriminate use of it, which is fashionable in certain schools of thought, does have its dangers. One should learn to recognize these, and strive to avoid them.

In considering fields of problems of whatever complexity – in which, for instance, random sets, functions, sequences of functions and so on can occur together – the most

---

6 Our preference for $Y(t)$, rather than the more usual $X(t)$ as a notation for a generic random function, depends mainly on the fact that an $X$ is often used as an 'ingredient' in the construction of $Y(t)$ At other times, $x$ is used as a variable in place of t, and, anyway, in the graphical representation it is always convenient to think of the ordinate as $y$, and the abscissa as $t$ or $x$.

7 Like, for instance, velocity $Y'(t)$ at an instant, measured with a speedometer; or the maximum or minimum of $Y(t)$ in an interval $(t', t'')$, measured with instruments like a Max–Min thermometer.

general way of interpreting and applying the concepts exhibited in Section 2.3.5 is always the same; that is the following.

One goes back to the finest possible partition into 'atomic' events – not themselves subdivisible for the purposes of the problem under consideration – and these are considered as *points* constituting the set $\mathscr{Q}$ of 'possible outcomes'. This abstract space is the 'space of alternatives', or the 'space of outcomes': in certain cases, such as the examples of Section 2.3, it may be convenient to think of it as embedded in a larger and more 'manageable' space, and to regard this latter as the 'space of alternatives'.

In this scheme of representation, each problem (by which we mean problem concerning the alternatives $\mathscr{Q}$) reduces to considering 'the *true* alternative' (or 'the one which will turn out to be verified', or however one wants to express it), as a *random point* in $\mathscr{S}$ or, if we wish to be precise, in $\mathscr{Q}$. Let us call this point $Q$: it expresses everything there is to be said. Were we to lump together in $\mathscr{S}$ all possible problems, this space would be the space of all possible histories of the universe (explained as far as the most unimaginably minute details), and $Q$ would be that point representing the true history of the universe (explained as far as the most unimaginably minute details).

Each event in this scheme is evidently interpretable as a set of points. $E$ is the set of all points $Q$ for which $E$ is true; for example, it is the set of all individual 'histories of the universe' in which $E$ turns out to be true. With the interpretation 1 = TRUE, 0 = FALSE, one could also say that $E$ is a function of the point $Q$ with values 1 on points $Q$ of the set $E$, and 0 elsewhere (the indicator[8] function of the set $E$).

Similarly, each random quantity is interpretable as a real-valued function of the points $Q{:}X = X(Q)$ is the value which $X$ assumes if the *true* point is $Q$. The preceding case, $E = E(Q)$, is simply the particular case which arises when the function can only take on the values 0 and 1.

The same is true for random entities of any other kind: for example, a random vector is a vector which is a function of the point $Q$.

2.4.2. That all this can be useful and convenient as a form of representation is beyond question. But things are useful if and only if we retain the freedom to make use of them when, and only when, they are useful, and only up to the point where they continue to be useful. A scheme that is too rigid, too definitely adopted and taken 'too seriously', ends up being employed without checking the extent to which it is useful and sensible, and risks becoming a *Procrustean bed*.

This is what happens to those who refer themselves too systematically to this scheme. Pushing the subdivision as far as the 'points' perhaps goes too far, but stopping it there creates a false and misleading dichotomy between the problems belonging, and not belonging, to the field under present consideration. The logical inconvenience which this already creates in the range of possibility will become far more dangerous and insidious when probabilities are introduced into such a structure.

An analogy between events and sets exists, but it is nothing more than an analogy. A set is effectively composed of elements (or points) and its subdivision into subsets

8 In a different terminology, the indicator function is also called the characteristic function: this term has many other meanings, and, in particular, in the calculus of probability it has a different and very important meaning for which it must be reserved (see Chapter 6).

necessarily stops when subdivision reaches its constituent points. With an event, however, it is possible, at all times, to pursue the subdivision (although in any application it is convenient to stop as soon as the subdivision is sufficient for the study in progress, otherwise things get unnecessarily complicated). The elements of the 'final subdivision' we have interpreted as 'points', but any idea which does not take into account the relative, arbitrary and provisional nature of such a delimiting of the subdivision, which thinks of it as 'indivisible', or as 'less subdivisible', or in any way different from all other events, is without foundation and misleading. For instance, it would be illusory to wish to distinguish between events corresponding to 'finite' or 'infinite' sets, or belonging to finite or infinite partitions, as if this had some intrinsic meaning. There is even less justification for retaining, as necessary, topological properties which happen to be meaningful in $\mathscr{S}$. The latter we referred to as 'space', instead of 'set', simply to use a more expressive language, and also because topological structures often exist and have interest in certain spaces by virtue of the nature of the spaces themselves, even when not required for any reason pertaining to the logical or probabilistic meaning.

2.4.3. Other objections, which we will develop a little more in the Appendix, would lead us to impugn even more radically the validity of the above representative scheme (and of many other things that we have hitherto admitted and which, for the moment, we continue to admit). As an example, we note the fact that all sets (or the 'points' of them) must be accepted as having the meaning of events.

In general terms, it will always be a question of examining if, and in which sense, a statement really constitutes an 'event', permitting, in a more or less realistic and acceptable form, and in a unique way, the 'verification' of whether it is 'true' or 'false'.

What should be said concerning statements that are 'verifiable' only by means of an infinite number of observations, or by waiting an infinite length of time, or by attaining an infinite precision? A critical attitude in this respect could lead one not to consider as 'events' the fact that $X$ has exactly the value $x$, or belongs to a set of measure zero (e.g. is rational), but only the fact that $X \in I$ for a set $I$ 'up to sets of measure zero' (and this, although it eliminates some difficulties, introduces others), or 'up to an error $<\delta$, that can be chosen as small as desired, but nonzero', and so on. Even more radical are the difficulties of 'complementarity', which appeared first in quantum physics but can be detected on a smaller scale in more everyday examples: $A$ and $B$ are events (observable), but it is not possible to observe both of them, and, therefore, it is not possible to call the product $AB$ an event (observable).

All this, in addition to the specific reasons already given in the main text (and to which we return in the next paragraph), reduces the value of the reduction to 'points'. Indeed, it is symptomatic that, precisely in connection with arguments of this kind, von Neumann developed a 'geometry without points' (in 'Continuous geometries', *Proc. Nat. Acad.*, **22** (1936), 92–100 and exemplified *Proc. Nat. Acad.*, **22** (1936), 101–108) where, as he says: 'The point which we wish to stress is that the investigations described above show an unbroken trend *away from the notion of the point*'. The studies to which he alludes are those of K. Menger and G. Bergmann (on linear spaces), of F. Klein, G. Birkhoff and O. Ore (on lattices), and discussions with J.W. Alexander and H. Veblen.

Even more strictly in accordance with the considerations in the text, appear to be the studies of St Ulam (in the 'von Neumann lecture', Princeton (1963), still unpublished), since he also refers himself to structures *open* to the adjunction of new entities as new circumstances arise. A 'continuous geometry' of von Neumann, on the other hand, is a closed structure, although very rich, containing linear systems of any dimension $c$, with $c$ any real number between 0 and 1 (the empty and complete systems, respectively). Ulam says: The indications are … that *there are no atoms of simplicity* and, which is most strange, one would almost be tempted to say that in the physical world the set-theory *axiom of Regularity* – that is to say, that *every set contains a minimal element with respect to the relation of "belonging to a set" – does not hold*!'.[9]

## 2.5 Logical and Arithmetic Operations

2.5.1. Having, through the convention 1 = TRUE, 0 = FALSE, given to events an interpretation that makes them particular random quantities, it becomes both possible and useful to take advantage of this unification in order to effect also an appropriate unification of the operations related to them. Usually, and inevitably, prior to such a convention,[10] one considers two distinct series of operations: the (Boolean) *logical operations*

$$\wedge\, logical\ product; \quad \vee\, logical\ sum; \quad \sim negation$$

applicable only to *events*; and the *arithmetic operations*

$$\cdot\, product; \quad +\, sum\, (\text{and their inverses} : / \text{ and} -)$$

applicable only to *numbers*.

We have already touched upon the utility of certain applications of the arithmetic operations to events, automatically possible by the above convention (see Section 2.3.4, and also allusions in Chapter 1). We are now able not only to develop this extension systematically, but also to obtain a complete unification by extending, in the opposite direction, the logical operations into the field of numbers.

In fact, in the field of (real) numbers, we make the definitions:

$$x \wedge y = \min(x, y), \quad x \vee y = \max(x, y), \ \sim x = 1 - x\, (= \tilde{x}).$$[11]

It is immediate that the definitions agree with those known in the field of events (that is, of the idempotent numbers 0 and 1), whereas, obviously, the usual properties (which it would be beneficial to interpret and understand through examples in each of the two cases), always hold both for numbers and events:

9 The italics are present in the original for the last three words only.
10 Which, as I later discovered, had already been adopted by von Neumann in 1932 in his treatment of quantum mechanics; Appendix, Section 9.
11 As usual, we agree to place the tilde for 'complementary to 1' above, instead of in front, when dealing with a single letter. The same convention – using a bar rather than a tilde – was adopted by L. Dubins and L.J. Savage, *How to Gamble if You Must*, McGraw-Hill (1965), p. 64, and found to be of frequent utility.

$$\left.\begin{aligned}\sim(x\wedge y)&=\tilde{x}\vee\tilde{y}\\ \sim(x\vee y)&=\tilde{x}\wedge\tilde{y}\end{aligned}\right\}\quad\text{(duality of }\wedge\text{ and }\vee\text{ with respect to complements),}$$

$$\left.\begin{aligned}x\wedge(y\vee z)&=(x\wedge y)\vee(x\wedge z)\\ x\vee(y\wedge z)&=(x\vee y)\wedge(x\vee z)\end{aligned}\right\}\quad\text{(distributivity between }\wedge\text{ and}\vee\text{),}$$

$$\left.\begin{aligned}x\wedge x&=x\\ x\vee x&=x\end{aligned}\right\}\quad\text{(idempotence for }\wedge\text{ and }\vee\text{)}$$

(in addition to the obvious commutative and associative properties of $\wedge$ and $\vee$).

2.5.2. *Operations on events*. By virtue of what has already been said, it is not a question of making new definitions, but only of applying the general definitions to the case of the values 0 and 1; it remains only to establish agreement with the usual meaning.

By the *logical product* of two (or more) events $A$, $B$, we mean the event which is true if and only if all the factors are, and therefore false if at least one is false. If the factors can only be 0 and 1, both the arithmetic product and the operation *min* ($\wedge$) obviously enjoy the property that the result is 1 if and only if all the factors are 1. Therefore, *in the field of events*, the two operations of arithmetic product and logical product coincide; thus, we could always refer simply to the *product* of two events, without danger of ambiguity, and write $E = AB$. The symbol $\wedge$ might be used for greater clarity only in complicated cases; for instance,

$$E=(X+Y\geqslant 54)\wedge(Z\geqslant Y+12),$$

where the events are conditions (on random quantities $X$, $Y$, $Z$ etc.), written as parentheses, and the fact that they are events and not numbers could be overlooked.

By the *negation* of an event $A$, we mean the event that is true if $A$ is false and vice versa; obviously, we have 'not $A$' $= \sim A = \tilde{A} = 1 - A$, because $\sim 1 = 1 - 1 = 0$, $\sim 0 = 1 - 0 = 1$.

By the *logical sum* of two (or more) events $A$, $B$, we mean the event that is true if at least one of the summands is true, and therefore false if and only if they are all false. To this corresponds the operation *max* ($\wedge$), which gives 1 if at least one summand is 1, and 0 if all summands are 0. It is also obvious and well known that, with respect to negation, the operation is dual to that of the product:

$$A\vee B=\sim(\tilde{A}\wedge\tilde{B}).$$

This follows also from the properties stated generally for $\tilde{x}\wedge\tilde{y}$.

This allows us to obtain an arithmetic expression for the logical sum: taking complements and expanding, we obtain

$$\mathrm{A}\vee B=1-(1-A)(1-B)=A+B-AB, \tag{2.1}$$

and, similarly,

$$\begin{aligned}\mathrm{A}\vee B\vee C&=1-(1-A)(1-B)(1-C)\\ &=A+B+C-AB-AC-BC+ABC.\end{aligned}$$

In general, for $n$ summands,

$$E_1 \vee E_2 \vee \ldots \vee E_n = \sum_i E_i - \sum_{ij} E_i E_j + \sum_{ijh} E_i E_j E_h - \ldots \pm E_1 E_2 \ldots E_n, \tag{2.2}$$

where the sums have to be taken over all the $n$ events $E_i$ over all the $\binom{n}{2}$ products two at a time, over all the $\binom{n}{3}$ products three at a time, and so on, with alternate signs, up to the last term which is the product of all $n$ events with + if $n$ is odd, − if $n$ is even.

The *arithmetic sum* of two (or more) events $A$, $B$, is not, in general, an event, but a random number expressing the *number of successes*. In particular, $A + B$ has either the value 0 (if they are both false), or 1 (if one is true and the other false), or 2 (if they are both true). In general, as in this case, the relation between logical sum and arithmetic sum is the following: *both have the value* 0 *if every summand happens to be false* (no successes), whereas, otherwise, if true summands (successes) exist and number 1, 2, 3,…, in general $m$, *the* (*arithmetic*) *sum is that number*, whereas the *logical sum* always takes the value *one*; that is, does not take into account multiplicity,

$$(\text{logical sum}) = 1 \wedge (\text{arithmetic sum}) \tag{2.3}$$

or, explictly,

$$E_1 \vee E_2 \vee \ldots \vee E_n = 1 \wedge (E_1 + E_2 + \ldots + E_n). \tag{2.3$'$}$$

The fact of having two distinct notions is not, therefore, inconvenient but, on the contrary, is an advantage because both have their raison d'être. We are still faced with the problem of eliminating the ambiguity of the terminology – since we do not wish to be obliged to say 'logical sum' or 'arithmetic sum' every time. For this purpose it is sufficient to adopt the natural convention of using *sum* for the arithmetic sum, and *event-sum* for the logical sum (because only this is an event).

2.5.3. We observe that the operations introduced induce, over the field of real numbers, the structure of a *lattice*, with the operation ~ which enjoys many properties of the complement (in the algebraic sense), but is not exactly such, except in the field of events (the numbers 0 and 1). There, in fact, we have $x \vee \tilde{x} = 1$ (because either $x$ or $\tilde{x}$ is 1, and the other 0), in addition to $x + \tilde{x} = 1$, which is also valid for any $x$.

In addition, we observe that the expressions in arithmetic form for $\sim x$, $x \wedge y$, $x \vee y$ coincide (in the field of events) with those of Stone, where the sum has to be taken 'mod 2', however, in order to obtain a Boolean ring.

The conventions adopted here do not give rise to algebraic properties of this kind but seem to be the most suitable for expressing, simply and naturally, many things which are otherwise difficult to express.

We will give examples at the end of this chapter (Section 2.11) in order not to interrupt the flow of the argument, and we will often use similar simplifications. It will be seen that it is not only a question of expressions concerning events or random quantities: for identical reasons, the same conventions meet requirements which also occur in other fields.

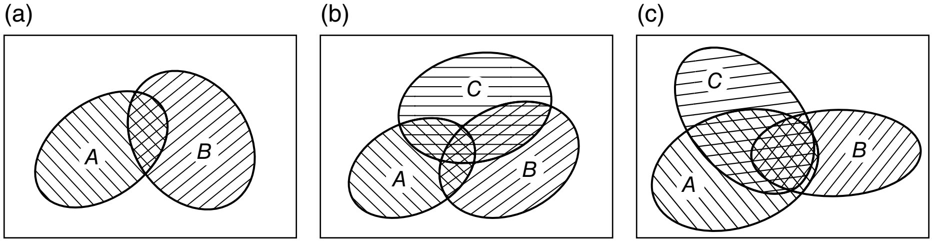



**Figure 2.1** Venn diagrams: the representations of events and their logical relationships in the set-theoretic interpretation: (a), (b) the cases of two and, respectively, three events with all (4 and, respectively, 8) constituents possible; (c) an example in which only six of the eight combinations give (possible) constituents.

2.5.4. We have mentioned, in Section 2.3, the set-theoretic interpretation. It is clear that, by interpreting the events as sets, the operations $\sim$, $\wedge$, $\vee$, which we have introduced, correspond in that context to the set-theoretic operations $\sim$, $\cap$, $\cup$ (*complementation*, *intersection*, *union*). For random quantities, understood as functions of the 'point' $Q$, $Z = X \vee Y$ is the function that, at each point $Q$, assumes the larger of the two values $X(Q)$ and $Y(Q)$:

$$Z(Q) = X(Q) \vee Y(Q) \quad \text{(and similarly for } \wedge\text{)}. \tag{2.4}$$

A geometrical representation (which is formally identical) is especially useful, particularly for didactic purposes, even if a genuine set-theoretic interpretation is lacking: it is that of the so-called 'Venn diagrams'. The events which one wishes to represent are drawn as areas of a rectangle, which itself represents the certain event. The areas are delimited with lines or, better, distinguished with different types of shading. In this way, one can illustrate visually the relationships that are supposed to exist among the different events: the existence, or not, of a certain intersection – distinguished by the overlapping of different shadings – the inclusion of one event in another; and so on. Of course, it is only in rather simple examples that clear figures, whose areas are not too contorted, are possible.

Shown in Figure 2.1a and 2.1b are the cases of two and three events, respectively, where all the four (or eight) intersections are nonempty, that is are possible events; whereas in Figure 2.1c two of the pair-wise intersections are not present.

## 2.6 Assertion, Implication; Incompatibility

2.6.1. We began this chapter by saying that, for You, every event, or proposition, can be either certain, or impossible, or possible. We then talked about possibility. The time has now come to translate these premises into a precise argument. We must make a distinction that, in the terminology proposed by B.O. Koopman,[12] could be called a distinction between *contemplated propositions* and *asserted propositions*. As considered so far, a proposition E is always a contemplated proposition (for which You, or anyone else,

12 *The Bases of Probability*, in Kyburg and Smokler, pp. 161–172.

could know whether it is true or false). Thus it remains, even if changed into $E = 1$, or $\sim E = 0$, or $(E = 1) = 1$, and so on, or, put into words, '$E$ is true', 'not-$E$ is false', 'it is true that $E$ is true'. Nothing is altered, because these are simply more or less extended ways of saying nothing more and nothing less than $E$.

To make an *assertion*, we have to step outside of the vicious circle by saying something extra-logical; such as 'I assert that $E$ is true', 'For You, it is certain that $E$ is impossible', 'For me, $E$ is possible': that is something expressing not a logical relationship between propositions, but a relationship between the proposition and the *speaker*.

To denote this succinctly, the symbol $\vdash$ has been introduced. If $E$ is a proposition, an event, then, by using $\vdash$ as a prefix, $\vdash E$ becomes the *assertion* that '$E$ is certain' (for someone). Naturally, $\vdash \sim E$ is the assertion that '$E$ is impossible', whereas by $\sim\vdash E$ we mean to denote the assertion that '$E$ is possible' (i.e. the nonassertion of both E and of not-$E$).

2.6.2. We shall not make much use of this symbol, because we think that, in general, the distinction will be clear from the context (for instance, by saying 'certainly'). It is useful, however, to draw attention to the importance of the distinction, and to illustrate the use of the symbol by giving some examples in order to fix all this in the reader's mind. In any case, these observations were necessary at this juncture in order to make it clear that certain expressions, which we will now introduce, *have to be taken as assertions*.

By saying that an event $A$ *implies* the event $B$, or that $A$ is *contained* in $B$, we mean to *assert* that $A$ cannot occur unless $B$ also occurs, or that $A\tilde{B}$ is impossible: in symbols $\vdash \sim A\tilde{B}$. Instead of $\sim A\tilde{B}$ one may also write $\tilde{A} \vee B$ or $A\tilde{B} = 0$, or $A \leqslant B$, or $B - A \geqslant 0$ (because the inequality is false only for $1 \leqslant 0$, i.e. for $A = 1$ and $B = 0$). It is always a question of ways of expressing $\sim A\tilde{B}$, independently of the fact that it is certain, or impossible, or possible, and these give assertions, simply by making the assertions. In order to write that '$A$ implies $B$' with the meaning, as we have said, of assertion, it will be necessary to write, for example, $\vdash A \leqslant B$. However, we will introduce some ad hoc symbols, to be understood as already having the value of assertions:

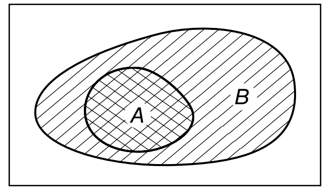
| | |
|---|---|
| $A \subseteq B. =. \vdash A \leqslant B,$ | $A$ implies $B$; |
| $A \equiv B. =. \vdash A = B,$ | $A$ is identical to $B$ (or $A \subseteq B \wedge B \subseteq A$), or, $A$ and $B$ are either both certainly true or both certainly false: (*certain*) *equality of A and B;* |
| $A \subset B. =. A \subseteq B \wedge \sim A \equiv B,$ | $A$ strictly implies $B$.[13] |

13 The *equality*, $A = B$, is the event that takes place if $A$ and $B$ are both true or both false, and this can happen for any $A$ and $B$ (except in the case of complementary events, $B = \tilde{A}$). However, in order not to make the language unnecessarily heavy, we will continue to say, as usual, 'equal', rather than 'certainly equal', and to write =, rather than ≡, except in ambiguous cases.

As regards *strict implication*, observe that it asserts that $A \leqslant B$ with certainty, but that $A = B$ is not certain. In other words, we exclude $A > B$, i.e. $A$ true and $B$ false, but we do not exclude the converse, $A < B$. Nothing is said concerning the possibility or impossibility of $A$ and $B$ being either both true, or both false. Observe that $A \subset B$ means $(A \subseteq B) \wedge \sim(A \equiv B)$ or $(\vdash A \leqslant B) \wedge (\sim\vdash A = B)$, which is very different from $\vdash \left[(A \leqslant B).\sim(A = B)\right] = \vdash A < B = \vdash \tilde{A}B$, which denotes the assertion that A is false and B is true.

The meaning of all these relationships is immediately, intuitively obvious under the set-theoretic interpretation.

**Figure 2.2** Venn diagram: the case of implication (inclusion).




2.6.3. The relationship of implication, which is clearly reflexive and transitive, induces, over any set of events, a partial ordering and in particular a lattice in which the operations $\wedge$ and $\vee$ (in the sense of 'maximal' element contained in those given, and 'minimal' element containing the given ones) coincide with those of logical product and logical sum, already introduced. This is evident above-all under the set-theoretic interpretation $A \subseteq B$ means that $A$ *is a subset* of $B$, possibly coincident with $B$ (this being excluded if we write $A \subset B$, affirmed if we write $A \equiv B$); hence the terms 'contains', 'is contained in', have the opposite meaning to 'imply', 'is implied by', instead of being synonymous as they might appear to be if one thought, for both terms, of the interpretation in terms of events.[14] In other words, in the Venn diagram for two events, which 'in general' (more precisely, for $A$ and $B$ logically independent) has the appearance of Figure 2.1, the part of $A$ not contained in $B$ must be missing (empty); in other words, $A$ must coincide with the doubly shaded area $AB$ (as in Figure 2.2).

If both regions with single shading are missing we have the case $A \equiv B$, and if the other two regions (double shading and no shading) are missing we have $A \equiv \tilde{B}$. Two other important cases correspond to the absence of the doubly shaded area (case of incompatibility: $AB \equiv 0$), or the absence of the nonshaded area (case of exhaustivity: $\tilde{A}\tilde{B} \equiv 0$).[15]

2.6.4. *Incompatibility*. By saying that two events $A$ and $B$ are *incompatible*, we mean to *assert* that it is impossible for them both to occur; i.e. that $AB$ is impossible: in symbols $\vdash \sim AB$. Instead of $\sim AB$ we can write $AB = 0$, or $\tilde{A} \vee \tilde{B}$, or $A + B = A \vee B$, or $A + B \leqslant 1$, or $A \leqslant \tilde{B}$, or $B \leqslant \tilde{A}$, always expressing the event $\sim AB$, independently of the fact that it is certain or impossible or possible. Each of these forms expresses the incompatibility; if it is asserted, we can write, e.g., $\vdash A + B \leqslant 1$, or $\vdash A \leqslant \tilde{B}$, which can be expressed, by reduction to the implication, as $A \subseteq \tilde{B}$. By saying that $n$ events $E_1, E_2, \ldots, E_n$ are incompatible, we mean to assert that they are pairwise incompatible ($\vdash E_iE_j = 0$, $i \neq j$); that is that at most one of them can occur. As a straightforward extension of $\vdash A + B \leqslant 1$, this can be expressed as $\vdash Y \leqslant 1$, where $Y = E_1 + E_2 + \ldots + E_n$ is the number of 'successes'; that is of the events $E_i$ that are true. The same definition also holds for an infinite number of events: in this case, instead of a non-negative integer, $Y$ could also be an infinite cardinal (e.g. that of denumerability, or of the continuum, or any other aleph). We note also that the condition $E_1 + E_2 + \ldots + E_n \equiv E_1 \vee E_2 \vee \ldots \vee E_n$, that is the coincidence of the logical and arithmetic sums,[16] is always characteristic of the case of incompatibility.

14 To avoid possible consequent mnemonic uncertainties about the meaning of $\subseteq$ (and hence the opposite meaning for $\supseteq$), it is sufficient to think of it as corresponding to $\leqslant$ ($\supseteq$ then corresponds to $\geqslant$), whose meaning is clear if we consider operations on the numbers 0 and 1 (events, indicator functions of sets).

15 The other cases are trivial: $A$ or $B$ or both would be determined, either certain or impossible.

16 For any non-negative (random) numbers the same conclusion is obviously valid: such equality holds if and only if at most one of them can be nonzero.

In other words, incompatible events are mutually exclusive; in the set-theoretic interpretation it is a question of *disjoint* sets, having an *empty intersection* (pairwise, and hence, *a fortiori*, for three or more).

2.6.5. *Exhaustivity.* By saying that two events $A$ and $B$ are *exhaustive*, we mean to *assert* that it is impossible for neither of them to occur; i.e. that $\tilde{A}\tilde{B}$ is impossible: in symbols $\vdash\sim \tilde{A}\tilde{B}$. Instead of $\sim \tilde{A}\tilde{B}$ one can (as above) write $\tilde{A}\tilde{B}=0$. or $\boldsymbol{A} \vee \boldsymbol{B}$, or $\tilde{A}+\tilde{B}=\tilde{A}\vee\tilde{B}$ (i.e. $2-(A+B)=1-AB$, $A+B=1+AB$), or $A+B\geqslant 1$, or $\tilde{A}\leqslant B$ or $\tilde{B}\leqslant A$; another form for the exhaustivity is therefore, for instance, $\vdash A+B\geqslant 1$. This lends itself easily to the extension of the definition to the case of $n$ events, or even to an infinite number. By saying that these are exhaustive (or, better, form an exhaustive family – but the phrase is cumbersome), we mean to assert that at least one of them must take place; that is, in the preceding notation, $\vdash Y\geqslant 1$. This shows the relationship between the two conditions. In the set-theoretic interpretation, it is a question of a family of sets which *covers* the whole set $Q$ of possible points (of course, there may be some overlapping); i.e. those sets of points for which the complement of the union is empty.[17]

## 2.7 Partitions; Constituents; Logical Dependence and Independence

2.7.1. *Partitions.* A *partition* is a family of *incompatible and exhaustive* events – that is for which it is *certain* that one and only one event occurs. The coexistence of the conditions $\vdash Y\leqslant 1$ and $\vdash Y\geqslant 1$ means, in fact, $\vdash Y=1$. A partition can be finite or infinite: partitions (and, for the simplest conclusions, in particular finite partitions) have a fundamental importance in the calculus of probability (which, as already indicated, will consist in distributing a unit 'mass' of probability among the different events of each partition).

It is, therefore, of importance to see now if, and how, one can reduce the general case, in which one considers any finite number of events $E_1, E_2, \ldots, E_n$, to that of a partition. We observe first of all that if, in particular, the $E_i$ are already incompatible, but not exhaustive, it will be sufficient to add on the extra event

$$E_0=1-\left(E_1+E_2+\ldots+E_n\right)\quad\left(\text{i.e., in another form, } E_0=\tilde{E}_1\tilde{E}_2\ldots\tilde{E}_n\right).$$

In the general case, we must consider the $2^n$ products $E_1'E_2'\ldots E'$ where each $E_i'$ is either $E_i$, or its complement $\tilde{E}_i$; formally, we can obtain them as the individual terms of the expansion $(E_1+\tilde{E}_1)(E_2+\tilde{E}_2)\ldots(E_n+\tilde{E}_n)$, which is identically 1, since each factor is 1. Some of the $2^n$ terms may turn out to be impossible and do not have to be considered: those which remain, and are therefore possible, are called the *constituents* $C_1, C_2, \ldots, C_s$ of the partition determined by $E_1, E_2, \ldots, E_n$, where $s\leqslant 2^n$.

17 Suppose that, instead of considering $Q$ – the space of possible points for You, now – one considers a larger space S which contains, in addition, certain points that are already known to be impossible (for instance, in the light of more recent information). In all the preceding cases, the statement that a set is *empty* must be replaced by *empty of possible points* – i.e. empty of points belonging to $Q$. In diagrams, one could think of the region $S\sim Q$ as drawn in black, and consider it as 'nonexistent'.

By observing that the given expansion has value 1, we have already established that we are dealing with a partition; on the other hand, the fact is evident per se (even more so under the set-theoretic interpretation). A partition is given by a family of disjoint sets that covers the space $Q$; or, in other words, into which $Q$ is subdivided – in the same way, for instance, in which Italy is divided into municipalities. If, instead, we perform any other division whatsoever, the partition given by the constituents is that into the 'pieces' resulting from such a subdivision. For instance, Italy east and west of the Monte Mario meridian, north and south of a given parallel, areas of altitude above and below 500 metres, areas more or less than 50 kilometres from the sea, belonging to a province the name of whose capital or main city begins with a vowel or consonant, and so on.[18]

Sometimes it will also be useful to introduce the (clumsy) notion of a '*multi-event*' for cases in which (provided we do not restrict ourselves to meaning 'event' in a purely technical sense) a partition might correctly be called an 'event with many alternatives'. Such is a game – a football game, for instance – with the three alternatives 'victory', 'draw' and 'defeat' (and possibly a fourth, 'not valid' because of postponement etc.). The same holds in the case of drawings from an urn containing balls of three or more different colours, for example 'white', 'red', 'black'; or throwing a die, or two dice, with possible points in the range 1–6, or 2–12, respectively. A multi-event with $m$ alternatives – more briefly an '$m$-*event*' – can always be thought of as a random quantity with $m$ possible values (e.g. 1, 2,…, $m$). In the case of a single die, the 'points' are precisely 1, 2,…, 6, whereas for the two dice it is irrelevant whether we use 2, 3,…, 12, or 1, 2,…, 11. The colours, or results of the game, could similarly be coded numerically. In speaking of an $m$-event we want, essentially, to emphasize the *qualitative* aspects of the alternatives. It is then appropriate to use the mathematical interpretation of them as unit vectors (1, 0,…, 0), (0, 1, 0,…,0),…, (0, 0, 0,…, 1) in an $m$-dimensional space. In this way, writing $E_h$ ($h = 1, 2,\ldots,m$) for the events[19] which consist in the occurrence of the $h$th alternative, an $m$-event can be identified with the random vector $(E_1, E_2,\ldots, E_m)$. The (arithmetic) sum of multi-events gives, therefore, the number of occurrences of the single results: for instance, $(W, R, B)$ = the number of drawings of White, Red and Black balls. We observe the analogy with the case of events, which could be handled in this same way, by substituting (1, 0) for 0 and (0, 1) for 1 (if the advantage of the symmetry seemed to compensate for the unnecessary introduction of the doubleton).

2.7.2. *Logical dependence and independence of events*. We define $n$ events (necessarily *possible*) to be *logically independent* when they give rise to $2^n$ possible constituents. This means that each of these events remains uncertain (possible) even after the outcomes of all the others, whatever they may be, are known: this explains the choice of terminology. In fact, let us suppose that one of the products is impossible, and therefore only a constituent in a formal sense – without loss of generality, take it to be $E_1E_2 \ldots E_n$. $E_1$ is possible, $E_1E_2$ may or may not be, and the same holds for $E_1E_2E_3$, $E_1E_2E_3E_4$, and so on. If one of these products is impossible, obviously all the subsequent ones are; the last one – the

18 Caution: do not think of separate parts of a unique nonconnected 'piece' as 'pieces' – the topology of the representation must be ignored. The 'piece' of Italy north-east of Monte Mario with altitude below 500 metres and more than 50 kilometres from the sea in a province beginning with a vowel is certainly composed of separated parts (for instance in the provinces of Ancona and Udine).
19 Necessarily incompatible and exhaustive.

product of all the $n$ events – is impossible by hypothesis, and therefore either it or one of the preceding ones must be the first to be impossible: suppose this is $E_1E_2E_3E_4$. This means that it is possible for events $E_1$, $E_2$ and $E_3$ to occur and that, knowing this, we are in a position to exclude the possibility that $E_4$ can be true. The events are therefore in this case – that is if the number of constituents is $s < 2^n$ – logically dependent.

Of course, if $n$ events are logically independent the subsets of the $n$ are, *a fortiori*, independent: the converse does not hold. Even if all their proper subsets exhibit logical independence $n$ events can still be logically dependent. As a simple example, let all the constituents in which the number of successes is even be possible, and no others; this imposes no restrictions on the result of any $n - 1$ events whatsoever, but for each event the result is determined once we know the results of the others.

2.7.3. If one wishes to consider more specifically the dependence of a *particular event* $E$ on certain others, $E_1, E_2, \ldots, E_n$, it becomes necessary to consider several cases. It is, in fact, possible that $E$ remains uncertain after we know the results of the $Ei$, whatever these may be: we then call it *logically independent*. On the other hand, it is possible that it will always be determined (either true or false), in which case we call it *logically dependent*. However, an intermediate case could also arise: the uncertainty or the determination of $E$ might depend on the actual results of the $E_i$; this we will call *logical semidependence*. We could be more precise and refer to logical semidependence *from below*, or *from above*, or *two-sided*, according to whether there exist outcomes for the $E_i$ which make *$E$ certain*, or *impossible*, or whether there exist outcomes of both types.

In order to characterize the various types of event, with respect to the fixed $E_i$, it suffices to consider the constituents determined by the $E_i$. We have $C_1 + C_2 + \ldots + C_s = 1$, and each event $E$ can, therefore, be decomposed into $E = EC_1 + EC_2 + \ldots + EC_s$. For any one of the summands, say $EC_h$, there are three possibilities: either $EC_h = C_h$ (if $C_h$ is contained in $E$), or $EC_h = 0$ (if $C_h$ is contained in $\tilde{E}$), or else $0 \subset EC_h \subset C_h$ (if both $EC_h$ and $\tilde{E}C_h$ are possible). The possible results for the $E_i$ correspond to the occurrence of one of the constituents $C_h$: according to whether $C_h$ is of the first, second or third type, $E$ turns out to be *certain*, *impossible* or remains *uncertain*, respectively.

The conclusions are obvious.

$E$ is *logically dependent* if constituents of the third type do not exist; that is if $E$ is a sum of constituents (of the first type). We could also say that $E$ is logically dependent on the $E_i$ if and only if it is expressible as a function of them by means of logical operations: in this case we have dependence by definition. The value (true or false) of such an expression is, in fact, determined by the values of the variables appearing in it; conversely, every such expression reduces to a *canonical form* as a sum of constituents and, therefore, the condition is also necessary. In this case, constituents of both the first and second types exist; otherwise, $E$ would have been either certain or impossible to begin with, contrary to hypothesis.

$E$ is *logically independent* if all the constituents are of the third type, and *logically semidependent* if some, but not all, are of the third type: in the latter case, we have semidependence *from below* if the others are all of the first type, *from above* if they are all of the second type, *two-sided* if there are some of each type.

If we consider the two events

$E'$ = the sum of all the constituents of the first type, and
$E''$ = the sum of all the constituents of the first and third types,

it clearly turns out that in each case $E' \subseteq E \subseteq E''$, and $E'$ and $E''$ are, respectively, the maximal event certainly contained in $E$, and the minimal event that certainly contains it – that is the events giving the best possible bounds.

We can then say that: $E$ is logically dependent if $E' = E''$ (and hence $= E$); logically independent if $E' \equiv 0$ and $E'' = 1$; semidependent from below, from above, or two-sided, if

$$0 \subset E' \subset E'' \equiv 1, \quad 0 \equiv E' \subset E'' \subset 1, \quad 0 \subset E' \subset E'' \subset 1,$$

respectively.

2.7.4. These notions of logical dependence and independence are meaningful more generally; they apply not only to the case of events, as considered so far, but also to partitions, or random quantities, or any random entities whatsoever. We will present the development for the case of random quantities, which is the most intuitive; it will then suffice to remark that the concept is always the same.

Random quantities (suppose, to fix ideas, that there are three: $X$, $Y$, $Z$) are said to be *logically independent* if there are no circumstances in which the knowledge of some of them can modify the uncertainty concerning the others. This means that if $X$, $Y$, $Z$ have, respectively, $r$ possible values $x_i$, $s$ possible values $y_j$, $t$ possible values $z_h$, then all the $rst$ triples $(x_i, y_j, z_h)$ are possible for $(X, Y, Z)$; that is the set $Q$ of possible points $(x, y, z)$ is the *Cartesian product of the sets*, $Q_x$, $Q_y$, $Q_z$ of possible values for $X$, $Y$, $Z$. In this form the definition is general: it is valid not only for $n$ (instead of 3), but also if the random quantities have an infinite number of possible values (for instance, those of an interval), or in the case of random entities of other kinds, or, generically, for partitions,[20] In other words, the condition means that nothing, no known interdependence, allows any further restriction of the set $Q$ of possible points over and above that resulting from the fact that the individual random quantities, or entities, must assume values in $Q_1, Q_2, \ldots, Q_n$.

2.7.5. The *logical dependence* of one (random) quantity on others (to fix ideas consider the dependence of $Z$ on $X$ and $Y$) has exactly the meaning that it has in analysis: $Z$ is a *function* (i.e. a *one-valued* function) of $X$ and $Y$, $Z = f(X, Y)$, the function $z = f(x, y)$ being defined for all the possible points $(x, y)$ of $(X, Y)$. *Logical independence* means that the set of possible values of $Z$ conditional on the knowledge of the values of $X$ and $Y$ (any pair of possible values $(x, y)$ for $(X, Y)$) is always the set of $Q$ of all the (unconditionally) possible values of $Z$. Intermediate cases, which are not worth listing in further detail, always give *logical semidependence*.

2.7.6. A critical observation is appropriate at this point, both as a refinement of the present argument and to exemplify various cases in which it is useful to examine whether the logic needs to be taken with a pinch of salt (see Appendix).

We will confine ourselves to a single example. Suppose that $X$ and $Y$ have as possible values all the numbers between 0 and 1, with the condition that $X + Y$ is irrational : then $Q$ is the unit square with an infinite number of 'scratches' removed – parallel to the diagonal and corresponding precisely to the lines $x + y =$ rational. For the partition into

20 A partition can be reduced to a random quantity by considering as such, for example, the index $i$ of $Ei{:}X = i$ if $Ei$ is true, provided the partition has at most the cardinality of the continuum, or is denumerable if we require integer values.

*points*, logical independence does not hold; it would hold, however, for every partition into vertical or horizontal stripes, however small.

Is it advantageous to say that we do not have logical independence when its failure is attributable to subtleties of this kind? Clearly, there is no categorical answer. It seems obvious that, depending on the problem and on one's intentions, one decides whether or not to take such subtleties into account (of course, one must be careful to be precise when such is required).

2.7.7. Finally, we make an observation which, strictly speaking, is unnecessary – being implicit in the very definition of 'possibility' – but which it is convenient to make and underline. All the notions we have encountered, or introduced – from incompatibility to logical dependence or independence – are relative to a given 'state of information'. They are valid for You (or for me, or him) according to the knowledge, or the ignorance, determining the uncertainty; that is the extent of the range of the *possible*, of the set $Q$ (yours, mine, his,...).

It is a question of relative and personal notions, but nonetheless objective, in the sense that they depend on what one knows, or does not know, and not on one's opinion concerning what one does not know, and what is, consequently, uncertain.

In order to avoid ambiguity, we must never forget that we are always speaking about uncertainty in the simple sense of ignorance. In particular, of course, we are dealing with matters traditionally attributed to 'chance' – a trace of this remains in the word 'random',[21] and in other expressions which we will be using. In general, however, we are concerned with any future matters whatsoever, and also of things in the past concerning which there is no information, or for which no information is available to You, or which You cannot remember exactly: we might even be concerned with tautologies. The various cases differ in one important aspect: that is the existence and degree of possibility and facility of obtaining, in one way or another, further information, should one wish to do so. This fact will, of course, be relevant in determining behaviour in decision problems, where it could be convenient to condition on the acquisition of new information. But apart from this, basically, it is convenient to regard any distinctions of this kind as unimportant. The only essential element, which determines and characterizes our object of study, is the existence of imperfect information – of whatever kind – and the situations of uncertainty in which, consequently, You might find yourself.

There is a prejudice that uncertainty and probability can only refer to future matters, since these are not 'determined' – in some metaphysical sense attributed to the facts themselves instead of to the ignorance of the person judging them. In this connection, it is useful to recall the following observation of E. Borel: 'One can bet on Heads or Tails while the coin, already tossed, is in the air, and its movement is completely determined; and one can also bet after the coin has fallen, with the sole proviso that one has not seen on which side it has come to rest'.

2.7.8. *Remark*. It might be useful to point out (or, for those who already know it, to recall the fact) that in the theory of probability one often uses the term 'independence' (without further qualification) to denote a different condition, that of *stochastic independence*, which refers to *probability* and will be introduced in Chapter 4.

---

21 *Translators' note*. The Italian word here is '*aleatorio*' (see French, aléatoire) from the Latin *alea* meaning *die*: 'alea jacta est !' – the die is cast !' – as Caesar said when crossing the Rubicon.

Be careful not to confuse it with *logical* independence – which we have just discussed – or with *linear* independence, which we will discuss in the next section. Both of these notions have an objective meaning; that is independent of the evaluation of the probabilities.

## 2.8 Representations in Linear form

2.8.1. *Basic notions*. When referring to the set $Q$ of possible 'points' in the case of two random quantities $X$ and $Y$, we tacitly interpreted the pair $(x, y)$ as Cartesian coordinates in the plane (which it was natural to take as the space of alternatives S). Similarly, for three or more points, we extend to ordinary space, or to spaces of any dimension (always in Cartesian coordinates).

This was simply a question of habit and, therefore, of convenience. One could have thought of any coordinate system; of a curved surface instead of a plane, or, in order to say more and in a better way, it is enough to think in terms of a space in a merely abstract sense, for which such distinctions of a geometric nature do not even make sense. With reference to the simplest case, it is sufficient that different pairs $(x, y)$ are made to correspond to distinct 'points'.

For further reasons, which we now wish to take into account – because, as we shall see in Chapter 3, they are essential for the theory of probability – it becomes important instead to think of S as a *linear* (*affine*) *space*. We shall call it the *linear ambit* and denote it by A because at times it will be convenient to consider as the space S not the whole of A but a less extensive manifold which contains $Q$. For example: if A is ordinary space, and $X$, $Y$, $Z$ are related by the equation $X^2 + Y^2 + Z^2 = R^2$, it might be convenient to think of S as the spherical surface on which one finds the possible points $Q$; these may consist of all the points of the surface, or a part of it, or just a few points, depending on other restrictions and circumstances and knowledge.

A representation that is *linear* with respect to certain random quantities (e.g. those considered initially) is such with respect to others that are linear combinations of them (but not with respect to the rest). If we require that linearity holds for the rest too, we have to extend the linear ambit A to new dimensions, as we shall see later.

The random quantities linearly represented in an ambit A themselves constitute a linear system, which we denote by S, and which is dual to A. One might ask whether it is useful to think of the two dual spaces, A and S, as superposed. In principle, the answer is no: in fact, only the affine notions have any meaning, and the metric, introduced surreptitiously by means of such a superposition, would be dependent on the arbitrary choice of the coordinate system that has to be superposed onto its dual. In general, for this reason, it is not even practically convenient. A unique exception is perhaps that of the case we considered first, in which we start from *events*, and it is 'natural' to represent them with unit, orthogonal vectors. In any case, whether or not this possibility is useful in a particular case, it is important never to forget that it is only the affine properties which make sense.

These properties also underlie the notions and methods fundamental to the theory of probability. On the other hand, the things in question are very elementary, and are currently applied without first introducing this formulation and terminology – which might well be considered excessively theoretical and, for the purpose in hand, disproportionately so. Nevertheless, if one is prepared to make the small effort necessary to picture the question in terms of the present scheme, many aspects of what follows will appear obvious and well-connected among themselves, instead of, as they might otherwise appear, unrelated and confused. So much so, that the preeminent – one might even say exclusive – rôle of linearity in the theory of probability has always remained very much in the background. This is, in part perhaps, because of the prominence given to the Boolean operations, and because of the nonimmediacy of the arithmetic operations on events when the latter are not identified with their 'indicators'. The present treatment is intended to provide the framework within which these observations will find their justification and clarification.

2.8.2. Let us begin by considering events $E_1, E_2, \ldots, E_n$, and, often, in order to be able to think in terms of ordinary space, we will, without essential loss of generality, take $n$ to be three.

The linear ambit A is the affine vector space in $n$ dimensions, with coordinate system $x_1, x_2, \ldots, x_n$, in which we will consider the values of the random quantities $X_1, X_2, \ldots, X_n$. In this case, the latter are the events $E_1, E_2, \ldots, E_n$, taking only the values 0 and 1: the set of 'possible' points consists at most, therefore, of the $2^n$ points (8, if $n = 3$) with coordinates either 0 or 1, and may be a subset of these. One sees immediately – as was inevitable – that the 'possible' points correspond to the $s$ ($s \leqslant 2^n$) constituents.

Given the special rôle of these points, it is convenient to think of the prism, of which they are the vertices, as a cube (or hypercube) and, therefore, to think of the Cartesian coordinate system $x_i$ as orthogonal and of unit length – with the reservation that this metric not be taken too 'seriously'.

The linear system L, of linear combinations of $E_1, E_2, \ldots E_n$, consists of random quantities $X = u_1E_1 + u_2E_2 + \ldots + u_nE_n$,[22] interpretable as the *gain* of someone who receives an amount $u_1$ if $E_1$ is true, plus an amount $u_2$ if $E_2$ is true, and so on (of course, the 'gains' may be positive or negative). The $X$ possess at most as many (distinct) possible values as there are constituents – namely $s$ – and the latter occurs if the corresponding 'possible points' are found on distinct hyperplanes $\Sigma_i u_i x_i = \text{constant}$.

An important example is that where $Y$ = *the number of successes*. In order to obtain this, it is sufficient to take all the $u_i = 1$ – a gain of 1 for each event – obtaining, as we have already shown directly, $Y = E_1 + E_2 + \ldots + E_n$. In this case, it is clearly not true that the possible points occur on distinct hyperplanes; if all the $2^n$ vertices of the hypercube are possible, they are, in fact, distributed over the $n + 1$ hyperplanes $Y = 0, 1, 2, \ldots, n$ according to the binomial coefficients $(1, n, \frac{1}{2}n(n-1), \ldots, n, 1), \binom{n}{h}$ being the number of possible ways of obtaining $h$ successes in $n$ events.

For the case $n = 3$, we shall denote the Cartesian coordinates of the ambit A in the usual manner, by $x, y, z$, and those of the dual system L by $u, v\ w$. If $X = uE_1 + vE_2 + wE_3$, then $ux + vy + wz$ is the value which $X$ would assume if $E_1$ takes the value $x$, $E_2$ the value $y$ and $E_3$ the value $z$. Given the meaning of the $E_i$ such values can only be either 0 or 1, and the value of the random quantity $X$ (e.g. gain) can only be one of those corresponding to the eight vertices of the cube (or to a part of it, if not all the vertices are possible).

22 In order to simplify this example we omit the constant $u_0$ (see Section 2.8.3).

Here are the coordinates of such vertices $Q$ together with the corresponding values of $X$:

$$\begin{array}{lcllllllll} Q & = & (0,0,0), & (0,0,1), & (0,1,0), & (1,0,0), & (0,1,1), & (1,0,1), & (1,1,0), & (1,1,1), \\ X & = & \ \ 0 & \ \ w & \ \ v & \ \ u & v+w & u+w & u+v & u+v+w. \end{array}$$

In particular, for $u = v = w = 1$, we see (as was obvious) that the number of successes is 0 in one case, 1 in three cases, 2 in three cases and 3 in one case. In addition (apart from the combinatorial meaning, $(1 + 1)^3 = 1 + 3 + 3 + 1 = 8$), this shows that, when projected onto a diagonal, the vertices of the cube fall as follows: one at each end, and three each at $\frac{1}{3}$ and $\frac{2}{3}$ of the way along the diagonal.

2.8.3. The sum $\sum_i u_i x_i$ (in particular $ux + vy + wz$) is a linear function both of $X$ (i.e. of its components $u_i$), and also of $\mathcal{Q}$ (i.e. of its coordinates $x_i$). We will denote it both by $X(Q)$ – thinking of it as ‘the value of a given $X$ as $Q$ varies’ – and also by $Q(X)$ – thinking of it as ‘the value assigned to different $X$ by the resultant $Q$’ The same operation, however, will still turn out to be useful independently of the fact that $Q$ is a possible point (i.e. $Q \in \mathcal{Q}$). That is, by replacing $Q$ by any $A$ in $\mathcal{A}$, writing $X(A)$ or $A(X)$:

$$A(X) = X(A) = \sum_i u_i x_i,$$

where the $u_i$ are the coordinates of the $X$ considered as points of $\mathcal{L}$, or, better, the components of $X$ considered as vectors of $\mathcal{L}$, and similarly the $x_i$ are coordinates (or components) of the $A$ considered as points (or vectors) of $\mathcal{A}$.[23] The expressions $A(X)$ or $X(A)$ then appear as *products* of vectors, $A$ and $X$, belonging to the two dual spaces $\mathcal{A}$ and $\mathcal{L}$.[24]

What we have said so far in this section is independent of the assumption that, rather than taking any random quantities whatsoever, we start with events, $X_i = E_i$ (as we did in Section 2.8.2, in order to fix ideas). Since it is convenient to consider not only the homogeneous linear combinations, $X = \sum_i u_i X_i$, as we have up until now, but also complete combinations with an additional constant, say $u_0$, we will always assume as added to the $X_i$ a fictitious random quantity $X_0$, taking the single value $X_0 \equiv 1$ *with certainty*. The summand $u_0 X_0$ has precisely the value $u_0$, with no alteration to the formula; we have only to take into account that there is an additional, fictitious, variable, $x_0$, and that, for all possible points (and, usually, also for every $A$ to be considered), we will have $x_0 = 1$.

2.8.4. *Linear dependence and independence*. We have considered $\sum_i u_i X_i$ $(i = 0,1,2,\ldots n)$, linear combinations (either homogeneous or complete) of $n$ random quantities $X_i$ ($i = 1, 2,\ldots, n$); $X$ is said to be linearly dependent on the $X_i$. It may be, however, that the $X_i$ are already linearly dependent themselves; that is that one of their linear combinations is identically zero (or constant: due to the inclusion of $X_0$ the two are essentially identical),

23 Given that the point $O$ (the origin) has meaning in both L and A, there is no risk of ambiguity in identifying points and vectors.

24 If one thinks of the two spaces as superposed – we have already said that, in general, this is not advisable – we would have the scalar product. In any case, one could write $AX$ and $XA$, instead of $A(X)$ and $X(A)$, thinking in terms of the product rather than writing it as a ‘function’. The main application, however, will be when $A = P$ (probability, prevision), and the omission of the parentheses in this case – although used by some authors – seems to give less emphasis to the structure of the formulae, and therefore to the meaning.

in which case at least one of the $X_i$ is a linear combination of the others and can be eliminated (because it already appears as a combination of the others). Geometrically, this means that the set $Q$ of possible points belongs to a linear subspace A′ of A, and hence it is sufficient to confine attention to A′: the extension from A′ to A is illusory – one adds only points which are certainly impossible.

We observe that linear dependence is a special case of logical dependence – that is that linear dependence is a more restrictive condition. Conversely, it goes without saying that logical independence is more restrictive than linear independence.

We now return, briefly, to the case of events, for even here the distinction between linear dependence and logical dependence is of fundamental importance for the theory of probability. The negation of $E$ depends linearly on $E$: in fact, $\tilde{E} = 1 - E$. On the other hand, the *logical product* $E = AB$, and the *logical sum* $E = A \vee B$, do not depend linearly on $A$ and $B$ (except when, under the assumption that $A$ and $B$ are incompatible, the logical sum has the form $A \vee B = A + B$). However, the logical sum does depend linearly on the two events and their product: $A \vee B = A + B - AB$. In general, the logical sum of three or more events depends linearly on the events themselves and on their products two at a time, three at a time,…, and finally the product of all of them (see Section 2.5.2). Apart from these cases of a general nature, however, it is possible that an event can be a linear combination of others 'by chance' (so to speak): an example can be found in Chapter 3, in connection with a probability problem, where an event $E$ is expressed linearly as a function of others by the following formula

$$E = \tfrac{1}{7}\left(3 - 2E_1 + E_2 - E_3 + 3E_4 + 5E_5 - 5E_6\right).$$

How can one tell whether or not such a linear dependence exists? It is sufficient to express all events as sums of constituents and then to see whether the matrix (consisting entirely of zeroes and ones) is zero or not.

2.8.5. The above considerations refer to the system $\mathscr{L}$, but linear dependence is still meaningful and important in the ambit $\mathscr{A}$. The interest there lies in considering the barycentre $P$ of two points $Q_1$ and $Q_2$ with 'masses' $q_1$ and $q_2$, where $q_1 + q_2 = 1$. By a well-known property in mechanics – which is, on the other hand, an immediate consequence of linearity – each linear function $X$ assumes at $P$ the value $X(P) = q_1X(Q_1) + q_2X(Q_2)$, and the same holds for the barycentre of three, or (leaving ordinary space) any number of points whatsoever. The property even holds if some of the masses are negative, but the cases in which we are normally interested are those with non-negative masses (usually, in fact, we will be dealing with probability).

The barycentre can, therefore, be any point[25] belonging to the *convex hull* of the points $Q_h$ under consideration. Consideration of the convex hull determined by the 'possible points', $Q \in \mathscr{Q}$ or, in other words, the convex hull of $\mathscr{Q}$ will play a fundamental rôle in the calculus of probability. Dually (and this property too, well-known and intuitive, will turn out to be meaningful in future applications), the convex hull is also the intersection of all

25 If the points $Q_h$ are infinite in number, then in order for this to be true we must also allow 'limit cases' of barycentres (which, in other respects, correspond to actual requirements of the calculus of probability, at least according to the version we will follow, in which we do not assume 'countable additivity'). Anyway, apart from questions of interpretation, this simply means that by convex hull we mean the set of barycentres completed by their possible adherent points.

the half-spaces containing $\mathscr{Q}$ In other words, if a point $P$ belongs to the convex hull $K(I)$ of a set $I$, then it is on the same side as $I$ with respect to any hyperplane not cutting the set – that is which leaves it all on the same side. On the other hand, if a point does not belong to the convex hull, there exists a hyperplane separating it from $I$ – that is which does not cut the latter and leaves it all on the opposite side with respect to the point. Translating all this into an analytic form: every non-negative linear function on $I$ is also such on $K(I)$; conversely, the property does not hold for any point not belonging to $K(I)$.

2.8.6. Returning to the case of the cube (Section 2.8.2), we already have a meaningful example, although a little too simple, of the way in which the convex hull varies as we consider all eight vertices or a subset of them (see Chapter 3, where the probabilistic meaning will also appear).

With this example in mind, it is now possible to make an observation which, although trivial in this context, is useful for explaining in an intuitive way our immediate intentions (Section 2.8.7) in cases where it could seem less obvious and perhaps strange.

In the space $\mathscr{A}$ we could represent the eight constituents by the vertices of the cube: we suppose that all eight actually exist, there is no need to consider other cases here. In the dual space $\mathscr{L}$, however, we could only represent the random quantities depending linearly on $E_1$,$E_2$,$E_3$. The eight constituents, considered as random quantities, could not be represented, and so neither could the random quantities derived from them linearly – unless these happened to be linearly dependent on the three fundamental events $E_i$. Does the method create a discrimination between events which have a representation as vectors in $\mathscr{L}$ and those which do not? If so, can we put the situation right?

The answer to the first question is no: the method creates no discrimination. The fact is that it enables us to consider more or fewer dimensions according to what we need. The representation in terms of the cube is sufficient for the separation of the eight constituents (as points of $\mathscr{A}$), and for the consideration of random quantities linearly dependent on the three $E_h$. If we wished, we could even reduce to a single dimension by considering only the random quantity $X = 4E_1 + 2E_2 + E_3$: this is sufficient to characterize the eight constituents, since $X$ can assume the values 0, 1, 2, 3, 4, 5, 6, 7. These values, incidentally, are obtained by reading the triple of coordinates as a binary number – for example (1, 0, 1) = 101 (binary) = 4 + 0 +1 = 5. If we were interested only in such an $X$ (up to linear transformations, $aX + b$), and in distinguishing the constituents, this would be sufficient. Similarly, if in addition to $X$ we are interested, for instance, in the number of successes $Y = E_1 + E_2 + E_3$, and nothing else, we could pass to two dimensions. Suppose, however, that, for reasons which depend on the linearization, we are interested in studying, in $\mathscr{L}$, either one of the constituents, or a linear combination of constituents not reducible to a linear combination of the $E_h$. In this case, it will be necessary to introduce a third dimension and then, if required, others…, up to seven. In general, if there are $s$ constituents we require $s - 1$ dimensions ($s$ if we include a fictitious one for the constant $X_0 \equiv 1$) *in order that everything geometrically representable in $\mathscr{A}$ is also linearly interpretable in $\mathscr{L}$.*

In fact, if in our case (that of the cube) we consider an eight-dimensional space, whose coordinates $x_h$ give the value of the constituents $C_h$, the possible points, $Q_h$, are the points with abscissa 1 on one of the eight axes (because one, and only one, of the eight constituents must occur). They are linearly independent in the seven-dimensional space $x_1 + x_2 + \ldots + x_8 = 1$: one of the $x_h$ is superfluous, but it makes no difference whether we leave it, or eliminate it and add a fictitious coordinate $x_0 \equiv 1$. In terms of $\mathscr{L}$, we can therefore obtain all the $X$ either as linear combinations $\sum u_h C_h$, for $h$ from 1 to 8, or for $h$ from 0 to 7 (excluding $C_8$ but adding the fictitious $C_0 \equiv X_0 \equiv 1$).

Conclusion: everything can be represented linearly provided one takes a sufficient number of dimensions. It is possible, and this provides a simplification, to reduce this number by projecting onto a subspace (although in this way we give up the possibility of distinguishing between those things which have the same projection). Thus, for instance, different possible cases may be confounded into a single one, or even if we take care to avoid it, barycentres arising from different distributions of mass may be confounded. In the case of the cube, for example, each internal point can be obtained as the barycentre of $\infty^{7-3} = \infty^4$ different distributions of mass on the eight vertices.

2.8.7. In the general case, considering any random quantities whatsoever, the same circumstance arises and has even greater interest. Suppose we consider the ambit $\mathscr{A}$ relative to $n$ random quantities $X_i$ ($i = 1, 2,\ldots, n$) and, for simplicity, let us assume that all the real values are possible and compatible for the $X_i$: that is that all the points of $\mathscr{A}$ are possible ($\mathscr{A} = \mathscr{Q}$). It follows that every random quantity $Z = f(X_1, X_2,\ldots, X_n)$ is *geometrically* individuated in $\mathscr{A}$ (to each point of $\mathscr{A}$ there corresponds, in a known way, a value of $Z$), but is not *vectorially* represented in $\mathscr{L}$ unless it is a *linear* function of the $X_i$. If such a vectorial representation for $Z$ is needed, however, it is sufficient to add on a new dimension for it – that is to introduce an extra axis, $z$, or, if one prefers, $x_{n+1}$, on which $Z$ can be represented.

To give an intuitive illustration: in the plane $(x, y)$ every function $z = f(x, y)$ already has a geometrical representation (visually through contour lines), but in order for $z$ to appear *linearly* in the representation it is necessary to introduce a new axis, $z$, and to transfer each contour line to the corresponding height, obtaining the surface $z = f(x, y)$.

As a practical example, in fact one which continuously finds application, an even simpler case will suffice. We have a single random quantity, $X$: by taking the $x$-axis as the ambit $\mathscr{A}$, we represent, by means of its points, all the possibilities (values $x$) which determine, together with $x$, every function of $x$, $f(x)$. However, if we are interested in the linear representation of a given $f(x)$ we must introduce a new axis, $y$, and on it represent $y = f(x)$. The linear ambit A will be the plane $(x, y)$, but for the space $\mathscr{S}$ we could more meaningfully consider the curve $y = f(x)$, whereas $\mathscr{Q}$ could be a set of points on such a curve (if not all values are possible for $X$). It will be, so to speak, the set $\mathscr{Q}$, previously thought of on the $x$-axis, projected onto the curve $y = f(x)$. We note, incidentally, that this illustrates the observation made in Section 2.8.1 regarding the nonidentification of A and S. The criterion which has been followed can be explained in the following way: we delimit S by taking into account the ‘essential’ circumstances, considering as such the fact of studying $X$ together with a given $Y = f(X)$, whatever the random quantity $X$ may be; we do not take into account the ‘secondary’ circumstances, considering as such the particular facts or knowledge which, in certain cases or at certain moments, lead us to exclude the possibility of $X$ attaining certain values.

The most important practical case (which we have already mentioned) is the simplest one: that of $X$ and $Y = f(X) = X^2$. The curve is the parabola $y = x^2$, and the linear system L consists of all the polynomials of second degree in $X$; $aX^2 + bX + c$. Suppose that we are interested in barycentres of possible points $Q_h$ with given masses $q_h$. If the points are taken on the parabola we obtain a point $\bar{x}, \bar{y}$, which is meaningful for both coordinates, whereas if we leave the points on the $x$-axis the barycentre would give the same $\bar{x}$, but no information about $\bar{y}$.

Obviously, if we were interested in considering $Z = X^3$ also (i.e. extending L to polynomials of the third degree) it would be necessary to take the space $(x, y, z)$ as the ambit A, the curve $y = x^2$, $z = x^3$ as the space S, and to project onto it the set $Q$ already given; and so on.

## 2.9 Means; Associative Means

2.9.1. Within this representation, we will take the opportunity to present, in an abstract form, a notion which has great practical and conceptual importance in all fields, and which, in what follows, will above all prove useful in connection with probabilistic and statistical interpretations. The notion in question is that of a mean. This is usually defined in terms of mere formal properties of particular cases, but (as Oscar Chisini pointed out) it has a well-defined and important meaning as a useful 'summary' or 'synthetic characteristic' of something more complicated.

A prime example (already considered in the preceding pages) is that of the barycentre, or, arithmetically, that of the arithmetic mean (in general weighted) of the coordinates of the point masses. It is well known how, in mechanics, for many aspects and consequences, everything proceeds *as if* the whole mass were concentrated at the barycentre. In the language of statistics (which we will encounter mainly in Chapters 11 and 12) one would say that knowledge of the barycentre (and of the mass) constitutes, for certain purposes, a *sufficient statistic* (i.e. an exhaustive summary). For other purposes, in mechanics, it is necessary to know in addition the moments of inertia, and the exhaustive summary is then the collection of these items of information *of first and second orders*. It is convenient to point out in advance that knowledge of the second-order characteristics will also play an important rôle in statistics and in the theory of probability. Above all, it gives a powerful tool for studying problems in a way that is often sufficiently exhaustive, although summary.

2.9.2. Let us now consider the definition of mean according to Chisini, which is based precisely on this concept of an exhaustive summary. In this way we impart to the notion the *relative functional* meaning conveyed by '*tailor-made*' (better the German *Zweckmässig*, whose equivalent is missing in other languages: zweck = purpose, mässig = adequate). According to Chisini,[26] '*$x$ is said to be the mean of $n$ numbers $x_1, x_2,\dots, x_n$, with respect to a problem in which a function of them $f(x_1, x_2,\dots, x_n)$ is of interest, if the function assumes the same value when all the $x_h$ are replaced by the mean value $x$: $f(x_1, x_2,\dots, x_n) = f(x, x,\dots, x)$*' Here we are considering the simplest case, without *weighting*, but the concept is still the same in the latter case, and in that – as we shall see in Chapter 6 – of *distributions*, even continuous ones.

2.9.3. The most important type of mean is the *associative* one. The defining property of associative means is that they are unchanged if some of the quantities are replaced by their mean (in the same way as, in order to find the barycentre, one can concentrate some of the masses at their barycentre). Independently, and almost simultaneously, Nagumo and Kolmogorov proved that the associative means are all, and only, the

26 O. Chisini, 'Sul concetto di media', in *Periodico di Matematiche* (1929); the topic is taken up again in an article by B. de Finetti in *Giorn. Ist. Ital. Attuari* (1931). The proof of the theorem of Nagumo and Kolmogorov can also be found there.

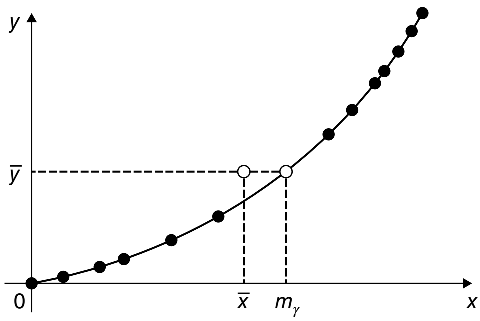



**Figure 2.3** Comparison between associative ($\gamma$-) means based on comparisons of the convexity of the functions $\gamma(x)$ used to construct them.

(increasing) transforms of the arithmetic mean. They are obtained by taking an increasing function $\gamma(x)$, and, given the values $x_h$ with respective weights $p_h$ $\left(\sum p_h = 1\right)$, instead of taking the barycentre, $\bar{x} = \sum_h p_h x_h$, one takes the barycentre of the corresponding $y_h = \gamma(x_h)$, $\bar{y} = \sum_h p_h y_h$ and then reverts to the 'scale' $x$ by means of the inverse function $m_\gamma = \gamma^{-1}(\bar{y})$ thus obtaining the $\gamma$-mean.

The procedure can be clearly 'seen' in the representation of the preceding paragraph. If we consider the example given there, we have $y = \gamma(x) = x^2$, and, of course, we must limit ourselves to the positive semi-axis in order that $\gamma$ be increasing.[27] It is a question of thinking of the masses $p_h$ as placed on the parabola; the barycentre is the point whose coordinates are $\bar{x}$ and $\bar{y}$, whereas $x = m_\gamma$ (obtained as shown in Figure 2.3) is the point to which corresponds (on the parabola) the same ordinate of the barycentre, the square root of the mean of the squares.

Considering the other function $z = x^3$ (either by itself in the plane $(x, z)$, or together with $y = x^2$ in the space $(x, y, z)$, as noted in Section 2.8.7), the barycentre would be $\bar{x}, \bar{z}$, respectively, $\bar{x}, \bar{y}, \bar{z}$, where $\bar{z}$ = the mean of the cubes $= \sum_h p_h x_h^3$, and $\sqrt[3]{\bar{z}}$ = the cube root of the mean of the cubes = the cubic mean of the values $x_h$ with weights $p_h$, and so on. In Chapter 6 we will say something about the most important associative means: these correspond to $\gamma(x)$ = powers (with any positive or negative real exponent whatsoever; if zero we have the limit case of the logarithm), and exponential. At this point, however, it is convenient to consider some general properties related to the notion of convexity of which we have spoken. This will also clarify a few questions which we will meet in Chapter 3.

2.9.4. The barycentre is always in the convex polyhedron (or, in general, the convex hull) determined by the point masses: in our example we can think of it both in the plane and in ordinary space. For the main conclusion of interest to us, the case of the plane is sufficient. If the masses are on a curve whose concavity is always in the same direction, or on a portion of the curve for which this is true, the barycentre is always in the area bounded by the concavity; hence: *the $\gamma$-mean is greater than the arithmetic mean if $\gamma$ (increasing) is concave upwards.* The quadratic mean is, therefore, greater than the arithmetic mean and so is the cubic mean: the question arises, can these two

27 Or the negative one. In fact, as is easily seen, $\gamma_1(x)$ and $\gamma_2(x)$ are equivalent with respect to the mean if (and only if) $\gamma_1 = a\gamma_2 + b (a \neq 0)$. If we change the sign of $a$ (i.e. change increasing into decreasing) nothing is altered. It is clear from the diagram, in fact, that a change in $y$, either of scale or sign, or a vertical translation of the curve, makes no difference.

be compared? Of course; it is sufficient to project the curve $y = x^2$, $z = x^3$ (explicitly, $z = y^{3/2}$) onto the plane $(y, z)$: the concavity is upwards and so the cubic mean is greater.

Even without the graphical comparison, it is sufficient to take into account that 'greater relative concavity' (in the sense that a diagram would display) corresponds, locally, to a greater value of $\gamma''(x)/\gamma'(x)$ (in the interval of interest if the function is not everywhere invertible). In the above example, we have $y''/y' = 2/2x = 1/x$, $z''/z' = 6x/3x^2 = 2/x$ and so, for $x > 0$, $z''/z'$ is always greater. More generally, since for the powers $\gamma(x) = x^c$ one has $\gamma''/\gamma' = c(c-1)x^{c-2}/cx^{c-1} = (c-1)/x$, the mean increases with the exponent; this also holds for $\log x$ (the limit case as $c \to 0$: $\log x \cong (x^c - 1)/c$): in fact, $\gamma''/\gamma' = -x^{-2}/x^{-1} = -1/x = (0-1)/x$. This particular choice ($c = 0$, $\gamma = \log$) gives the geometric mean, which, in the case of two, or more generally $n$, values with equal weights (the 'simple', unweighted case) assumes the more familiar forms: $\sqrt{(x_1x_2)}, \sqrt[n]{(x_1x_2\ldots x_n)}$, respectively. For $c = 1$, we have the harmonic mean, the reciprocal of the mean of the reciprocals.

From the fact that $-1 < 0 < 1 < 2 < 3$ it follows that in the above-mentioned cases, for example, we have:

harmonic < geometric < arithmetic < quadratic < cubic.

2.9.5. *Remarks.* Although it may seem strange to do so, we conclude by saying that the following observation is important: the barycentre of points which are on a curve (other than a straight line) is not a point on the curve – unless perhaps 'by chance'. In the same way, the barycentre of points on a surface (not a plane) is not, generally speaking, a point of the surface; and so on, in any dimension. The observation may seem strange because it is so obvious: its obviousness, however, results from the demonstration in terms of the above representation. How many people would recognize the fact before having their attention drawn to it? In facing real problems one often reasons as if what one considers strange, and even absurd, is precisely this fact!

## 2.10 Examples and Clarifications

2.10.1. Examples are always useful in order to give a sense of concreteness to concepts introduced in a general and abstract form. In this case, they will serve in addition to underline the meaning and importance of certain refinements, either already mentioned in passing or to be added soon, and also to introduce, before we yet talk about probability, a few of the kinds of situation which we will repeatedly come across in various problems.

Above all, by selecting widely differing examples we intend to remove any possible residual doubts that might lead to restrictive interpretations of the field of uncertainty to which we refer ourselves. The subject matter to which the uncertainty refers is irrelevant: political or economic events, meteorological phenomena, historical or scientific conjectures, judicial investigation, personal or everyday affairs, competition in sport, or any other field in which uncertainty and imperfect knowledge are present. This includes of course –and they in no way differ from the others – the traditional games of chance. This latter is, in fact, the least interesting case, because it leads to a standardized scheme in which all the conceptual and substantial aspects of the problem are made to disappear.

2.10.2. *Examples of events.* Will a given candidate, on a given occasion, succeed in getting elected (for instance, as a senator, a mayor, a member of a committee, a president of a society or of the university), or in passing (for instance, a student sitting an examination), or in being the winner (in a contest or a lottery, at Bingo, in a sports competition, in a game of cards or chess, or anything else) and so on? Will a vote turn out to be favourable – for instance, for a given law, or for an issue of confidence facing a government and so on? Is the accused in a given trial really the murderer? And, in any case, will he be convicted as such? Is the approaching tram the one I am waiting for? Will the next child of a given couple be a boy? Will it rain tomorrow at a given place? Will the next attempt at a soft landing on the moon be successful?

In all cases, and in various ways, if we want to be more detailed, or to extend the questions, we often conveniently express ourselves in terms of random quantities. In the examples of elections and voting, we might ask the following sorts of questions. How many votes are favourable? How many against, invalid or abstentions? What is the percentage of those in favour? In the case of examinations, contests and competitions, what is the mark or position obtained? And when – what year, day or moment – will the event in question occur (moon launch, trial verdict, vote, birth of the particular baby etc.)? Or, alternatively, how many will succeed – among those participating in an examination, contest, sports event and so on? Which one among them – identified by entry number, or position in alphabetical order – will attain first place, or second place? Who, within a given age limit, will be best placed? In a competition with several stages, or legs, who among the entrants will maintain, or improve, or worsen, their position with respect to the previous placings?

In other cases one uses different terminology. Random point: for instance, the point of the lunar surface which will next be reached. Random set: the set of those who pass an examination, the set of points on the earth's surface on which rain will fall tomorrow, the set of instants at which the temperature at a given place is below, above, or at, zero. Random function: the temperature at the above-mentioned place, the score during a competition, the number of votes of confidence since a certain date and so on, all considered as functions of time. If one wishes to avoid reference to irrelevant items of information (for instance, by referring to an entry number rather than to the individual concerned) it is preferable to speak of a multi-event, rather than a random quantity, and so on.

2.10.3. It is clear that in all cases it would be possible to go into more and more detail, and if all the cases we have mentioned were considered simultaneously we would arrive at even more minute subdivisions. And to these cases could be added others, ad infinitum. To arrive at a final subdivision into 'points' – not further divisible – would at least imply the construction of all possible 'histories of the universe', distinct in every detail. These would include, for example, the precise specification, instant by instant, of the position of every atom, and of the thoughts and moods of each individual – including, possibly, beings, more or less similar, living on other worlds. Even if we limit ourselves to much more restricted problems, an exhaustive description, though very much reduced in scale, would by no means turn out to be more realistic. Consider a single toss of a coin: unimaginable faculties would be needed if we wished to provide a description, with such absolute precision, of a single one of the possible ways in which a person tosses the coin, the air influences the movement, and every peculiarity of the ground and of the coin at the point and position of the latter's fall gives rise to successive movements, and so on, until the coin comes to rest. But this would still be nothing, because, instead, we must imagine and distinguish the totality of such ways.

We have pushed ourselves to absurd lengths – in a way pointless in itself – but perhaps this will serve to illustrate the thesis that it is inappropriate to distinguish between events represented by 'points', or by 'sets', thinking of it as something systematic, rather than being dependent on momentary conveniences of representation.

2.10.4. This has been said to emphasize the considerations already made (in Section 2.7.7 and elsewhere), but it is even more necessary to underline the sense in which an event (random quantity etc.) has to be – as we said – something 'well determined'. This means that the formulation must be unambiguous and complete, in such a way as to rule out any possibility of argument (for instance in the case of a bet which is based on it). To give an example: 'A.N. Other wins the lottery'[28] is an event only if the person A.N. Other, of whom we are speaking, is perfectly individuated, along with the circumstances that make the statement precise. Examples of the latter might be: win in next week's drawing; or in the first week that he plays; or any week of this year; and so on. It should also be made precise, or understood, whether possible wins in partnership with others are to be included, or not, along with any other possible aspects allowing ambiguity. By changing the individual, or any of the circumstances or provisos, we obtain other events, all different from each other. We say this only to avoid the situation where, being familiar with other terminologies, someone might think that they should be called 'identical events' or, even worse, 'trials' of 'the same event', which consists in 'winning the lottery'.

Conversely, two events expressed in completely different ways are identical – that is they are the same event – if we know that the occurrence of either one of them implies the occurrence of the other. Suppose, for instance, that we know for certain that this week A.N. Other is going to play the 'straight' three numbers 21–63–82 on the Roman wheel, and nothing else: in this case, the two events 'A.N. Other is going to win the lottery this week' and 'This week the numbers 21, 63 and 82 will come out on the Roman wheel' are identical. On the other hand, in order to demonstrate that it would be wrong to think in terms of the identification of a 'fact', we note the following: 'A.N. Other is going to win the next time he plays' and 'Next week the youngest person playing is going to win' are two *distinct events* which might, by chance, turn out to be the *same fact* if next week A.N. Other plays and wins and, in addition, happens to be the youngest player. This example also serves the purpose of making clear that there is no need to identify explicitly the person and the drawing (either by the date or, possibly, the wheel) so long as, by some means or other, it turns out that whether we must call the *statement* true or false is well determined.

2.10.5. One could object, with reason, that such a requirement is practically unrealizable, and that, in fact, it is not even realized in the example which we have just given. For instance, how is the statement 'A.N. Other is going to win the lottery the next time he plays' to be evaluated if A.N. Other never plays again for the rest of his life? This should be made clear by means of some arbitrary convention. In most cases of this kind, however, we shall interpret the statement in a sense which falls outside the present concept of an event, but which leads to a generalization (conditional event) that we will consider

28 *Translators' note*. Every Saturday in Italy, at each of ten cities, a drawing takes place of 5 from 90 possible numbers. To enter the lottery, one places a bet, prior to the drawing, specifying which combination(s) of numbers (up to a maximum of 5) one thinks will be drawn in a chosen city. The device which produces the numbers is known as a 'wheel'.

explicitly later on (in Chapter 4). In addition to being *true* (=1) or *false* (=0) it could also be *void* (= Ø). In terms of a bet, this means that it could not only result in gain or loss, but could also, in certain cases, be called off. If these things are not made clear explicitly, in a systematic way, then even the statement 'A.N. Other is going to win the lottery next week' might appear ambiguous, because of the doubt as to whether we mean 'false' or 'void' if A.N. Other does not play: in such a case one implicitly assumes certain refinements, but without justification. We will not labour this point, postponing further discussion until the appropriate place. In the same way, we do not enter into discussion of certain other questions, like perhaps the preceding ones, which may appear sterile but which, if misunderstood, give rise to numerous possible ambiguities and errors. In contrast to the above, these questions can be put off until later. Let us merely remark – in order not to seem mysterious – that, above all, it is a question of discussing the actual possibility of obtaining, within a given time and with greater or lesser certainty and precision, information concerning the events and quantities of interest, about which we are at present uncertain.

2.10.6. We now return to the examples that we considered before in order to draw attention to some of the kinds of problems that we will frequently meet in the future, and which will serve, for the time being, to illustrate the notions introduced in the preceding paragraph.

When we ask how many of the participants in an examination will succeed in passing, we have an example of a problem concerning the *number of successes*, $Y = E_1 + E_2 + \ldots + E_n$, where $E_h$ = 'the success of participant $h$' or, alternatively, one concerning the frequency, or percentage, of successes, $Y/n$. Other examples, chosen from the infinite number of possibilities, might include the following: the number of 'white balls in $n$ given drawings from an urn'; or of 'males among the first $n$ births registered in Orvieto next year'; or of 'those among the $n$ participants in a competition with many stages who maintain, after a given stage, their previous position'.

Clearly, $Y$ can only assume the values $0, 1, 2, \ldots, n$, and, obviously, these will all be actually possible if the events $E_h$ are logically independent. This means that the set of all those who pass an examination can, in fact, be any one of the $2^n$ subsets of candidates (including the whole set and the empty set); that is for each $h = 0, 1, 2, \ldots, n$, all the $\binom{n}{h}$ subsets of $h$ individuals are subsets for which $Y = h$. In the cases of examinations, drawings from an urn, births and so on, this will be true under most of the usual assumptions (and we shall see what these are shortly, when we turn to counterexamples). For the time being, however, we note that the $n + 1$ values can all be possible, even in cases where logical independence does not hold. Suppose, for instance, that $E_h$ means that 'the person placed in the $h$th position in a competition has reached some minimum prescribed score' (or time in a race, distance with a throw, height with a jump). It is possible that all, or none, or any intermediate number $h$, will succeed; in the latter case these are obviously the first $h$ and no others. We do not have logical independence since if $E_h$ is true, all the preceding ones are necessarily true, and if false all the following ones are false.

At the other extreme, it is possible that $Y$ is certain. This is the case, for instance, if $E_1$, $E_2, \ldots, E_n$ represent the drawing of white balls in $n$ successive drawings *without replacement* from $n$ balls, $h$ of which are white; then we certainly have $Y = h$, the number $h$ being known with certainty at the present moment. But in every case (drawings with replacement, examinations, sex of births) we find ourselves in the same situation if we are acquainted with the outcome as a whole, even though ignorant of the results of single drawings and so on. It is important to notice that the $E_h$ are, in this case, not only logically, but also *linearly, dependent* ($E_1 + E_2 + \ldots + E_n = Y = h$). The logical dependence assumes a concrete form in the fact that once all the white balls (or all the others) are out, the result of the subsequent drawings is certain (in any case, this is always so for the last drawing at least).

All intermediate hypotheses can be shown to be possible by the use of examples of a more or less artificial nature. The actual possibility of all $n + 1$ values is also compatible with linear dependence: if $n \geqslant 3$ we could have $E_1 = E_2$ with certainty if one thinks, for example, of the first two balls being drawn from an urn containing balls of the same colour in pairs. Restrictions on $Y$ may exist in the case of competitive examinations with a maximum number of awards available, or in the case of drawings without replacement of $n$ balls from an urn containing $N$ balls of which $H$ are white : in this case $n - (N - H) \leqslant Y \leqslant H$, the restrictions being real if the limits are $>0$ and $<n$, respectively.

An important case, of some interest since it is rather less obvious, is that in which all values except $n - 1$ are possible. We meet it in the example of 'maintaining rank in a classification', which, more abstractly, consists in considering the elements that remain fixed under a permutation. One of the many well-known different interpretations is the following: we put, more or less haphazardly, $n$ letters into $n$ envelopes, and we consider the random quantity $Y$ which denotes the number of letters correctly placed. Clearly, all outcomes are possible, except that of making just one error: one letter cannot be misplaced if all the others are in their own envelopes, since only the correct envelope then remains.

2.10.7. In the case of three or more alternatives (for each of $n$ multi-events[29]) we must consider for each of them the number of successes or realizations: for instance, $X$, $Y$, $Z$, with $X + Y + Z = n$, $X$, $Y$, $Z$ being the number of votes for, against or abstentions, out of $n$ votes; or wins, draws and losses out of $n$ games; or of bachelors, married men or widowers out of $n$ males; and so on, and so forth. Similarly for cases involving many alternatives: for example $X_1 + X_2 + \ldots + X_6 = n$ for occurrences of the points 1, 2,…, 6 when we throw $n$ dice, or a single die $n$ times (or, in the previous example, if we distinguish marital status and sex).

Problems of this kind are called problems of *subdivisions*: here we have been dealing with the subdivisions of the integer $n$ into a given number of (non-negative) integer summands, but more generally we could consider subdivisions of a given quantity $q$ into any kinds of summands whatsoever – non-negative real values $X_1 + X_2 + \ldots + X_m = q$. We often prefer to take $q = 1$, that is to reduce to percentages: in the preceding case we could also divide the numbers of occurrences by $n$, obtaining in this way the frequencies. A classical example is the subdivision of an interval (into $m$ parts with $m - 1$ division points). One could also imagine, however, the masses of the $m$ parts into which an object of mass $q$ breaks on falling; or, alternatively, the masses of $m$ materials from which it is constructed (for example $m$ metals if we are dealing with an alloy). We shall meet these kinds of problems again.

It is of interest to note that in such cases the $m$ random quantities are linearly dependent. Other quantities that have to be considered in connection with questions of this nature

29 Of course, this is also valid in the case of only two alternatives; in this case, however, it is trivial to take into account the number of occurrences of each of them since $Y = n - X$.

are also linearly dependent if they are linear combinations of them. As examples, we note the difference between votes for and against, or the total number of 'points' scored (taking 2 for a win, 1 for a draw). On the other hand, this would not be true, for example, for ratios, such as votes in favour divided by votes against, where one would have logical but not linear dependence.

2.10.8. In the above example of a ratio ($Z = Y/X$), and in others that will follow, the logical dependence will be functional dependence (in the clearest case, with $f(X_1, X_2, \dots, X_n)$ such that each $X_h$ turns out to be uniquely determined within the permitted field). Naturally, the given definition does not imply anything of this kind. Not only may the uniqueness fail – as when we consider points on the spherical surface $X^2 + Y^2 + Z^2 = 1$ with admissible values not constrained to be non-negative – but one might also consider all points of the sphere as possible (by substituting $\leqslant$ for $=$) without destroying the logical dependence. To see this, note that, given $X = x$ and $Y = y$, the possible values of $Z$ lie in the segment between $\pm\sqrt{(1 - x^2 - y^2)}$, which is a function of $x$ and $y$. Given that $X$, $Y$, $Z$ can all assume values between $\pm 1$, we have logical independence only for the case in which all the points of the cube $-1 \leqslant x, y, z \leqslant +1$ are possible: the exclusion of a single point, for example the origin, is sufficient to give logical dependence (to avoid it, we would have to exclude the points on the coordinate planes; i.e. the value 0 for each random quantity separately). One also has logical dependence if one excludes from the cube the points for which, for example, $X + Y + Z$ (or $XYZ$, $XY/Z$, etc.) is rational, or transcendental, or whatever (to avoid it, one should instead exclude separately, $X$, for example, being rational, $Y$ being transcendental, $Z$ being zero).

2.10.9. A case of logical dependence, which is of practical importance and frequent occurrence, is the following: given a number of random quantities, say $X$, $Y$, $Z$, we denote, by definition, the smallest of these by $X$, the middle one by $Y$, the greatest by $Z$. In this case, we exclude all those points which are not included in the dihedron $y - x \geqslant 0$, $z - y \geqslant 0$, even if the coordinates of the points are possible values for $X$, $Y$, $Z$ (unless all the possible values for $X$ are less than all the possible values for $Y$, and these are less than all the possible values for $Z$, in which case $X \leqslant Y \leqslant Z$ does not constitute a restriction). It is necessary to pay attention to circumstances of this kind, as the necessity of establishing and taking appropriate account of them could be overlooked.

If we take the example of a subdivision resulting from the splitting of a fallen object – let us say into three pieces, $X$, $Y$, $Z$ – the situation differs according to whether the criterion by which we rank them is the order of magnitude, or something else not depending on it. For instance, we might take the angle formed between the half-line starting from the point of fall and passing through the barycentre of the piece in question and the direction North, the angle being taken in a counterclockwise direction.

The same thing holds in the example we are about to consider now, where $X$, $Y$, $Z$ are the sides of a random prism (rectangle): for example, a block of stone, a building, a suitcase. We may or may not have more or less 'natural' circumstances which lead us to define, in each case, what we mean by 'length' ($X$), 'breadth' ($Y$) and 'height' ($Z$). Without getting bogged down in an analysis, which everyone can provide for themselves anyway, the answer seems to be easy for the suitcase, not always such for the building – the distinction between length and breadth may not be clear if there is no recognizable façade – and indeterminate for the block (unless we use conventions based on how it is temporarily situated with respect to North, East and the zenith). If we agree to call the maximum side the length, and the minimum side the height, we are in the other situation.

Given this random prism – and however we think of the problem, with the sides $X$, $Y$, $Z$ logically independent or not – let us consider its diagonal $U$, area $V$ and volume $W$. In either case, these are random quantities that are logically (and even, in a unique way, functionally) dependent on the preceding ones: $U = \sqrt{(X^2 + Y^2 + Z^2)}$, $V = 2(XY+ XZ + YZ)$, $W = XYZ$. Clearly, however, the dependence is not *linear*; when we return to the question, in Chapter 3, this example will serve to clarify, in an appropriate way, how, and why, certain reasonings about uncertainty, though seemingly obvious, are correct in some cases, but not in others (and this according to whether one has linear dependence or not).

## 2.11 Concerning Certain Conventions of Notation

2.11.1. As we announced in Section 2.5.3, and briefly mentioned in Chapter 1 (1.9.3 and 1.9.4), we will demonstrate, by means of examples, the utility that can be derived in many cases from the use of conventions introduced in the present chapter for simplifying the notation. To be explicit:

- the identification of TRUE and FALSE with 1 and 0;
- the 'lattice' operations for numbers.

2.11.2. The convention TRUE = 1 and FALSE = 0 turns out to be very useful also when applied outside of the field of events, to propositions or any 'conditions' whatsoever

*Examples.* $(x \geqslant a)$ is the function which = 0 for $x < a$ and = 1 for $x \geqslant a$; we could write such a function as $F(x) = (x \geqslant a)$, and, more generally,

$$F(x) = \sum_h p_h (x \geqslant a_h)$$

is the step-function with jumps $p_h$ at the points $x = a_h$; assuming that the $a_h$ are in increasing order of magnitude, this could also be written

$$F(x) = \sum_h c_h (a_h \leqslant x < a_{h+1}),$$

which denotes that in the given interval the value is

$$c_h = \sum_i p_i (i \leqslant h) = \sum_{i=1}^{h} p_i.$$ [30]

In the last example we used the function $(a \leqslant x < b)$, which is = 1 in the given interval and = 0 outside: more generally, we use $(x \in I)$ to denote the indicator function of the set $I$ (the function which = 1 if $x$ is in $I$, and = 0 otherwise).

30 Given the purely illustrative purpose of these forms of notation, we omit all the possible refinements that should be added, case by case, in specific applications: for instance, here, hypotheses of convergence if we are dealing with series: in an opposite sense the convention $a_{n+1} = \infty$ if $a_n$ is the last term, etc. The notation $\leqslant$ instead of $<$ etc. will vary from case to case.

Using such a function as a multiplier, one obtains immediately the restriction of a function to a given interval or set; for example,

$$x^2(x \geq 0) = 0 \text{ for } x \leq 0 \text{ and} = \mathrm{x}^2 \ \textit{for } x \geq 0;$$

$f(x) = x(1-x)(-1 \leqslant x \leqslant 1) = x(1-x)(|x| \leqslant 1)$ is equal to $x(1-x)$ for $x$ in $[-1, 1]$, 0 otherwise; and, more generally, for a function with a different expression in different intervals, for example,

$$f(x) = a(x+3)^3(-3 \leqslant x < -1) + (b - cx^2)(-1 \leqslant x < 1) \\ + a(3-x)^3(1 \leqslant x < 3),$$

or even (for a large, or infinite, number of intervals)

$$f(x) = \sum_h f_h(x)(h \leqslant x < h+1), \text{ or } f(x) = \sum_h f_h(x)(a_h \leqslant x < a_{h+1}).$$

*Remarks*. The examples in which the functions are denoted by $F(x)$ and $f(x)$, respectively, can be interpreted as the *distribution* function ($F$) and the *density* function ($f = F'$) of a distribution. These notions may already be familiar but will, in any case, be introduced in Chapter 6.

2.11.3. In the previous cases of summation, we have already seen the expression of the condition functioning as a multiplier in order to define each single sum-function, or (under another equivalent interpretation) specifying, for a given $x$, which terms had to be summed. The systematic usage of such a convention to this end, even in the absence of a useful interpretation in the first sense, would seem to be very convenient, both for clarity and typographical convenience. It replaces, in an advantageous manner, either explanations in the text, or complicated instructions to be composed under the summation (or integral) sign and so on.

The meaning of the following examples is self-evident:

$$\sum a_h(h \in H), \ \sum a_h(b_h \in B), \ \sum a_h(h \neq 0), \ \sum a_{hk}(h \neq k), \\ \sum a_{hk}(h \leqslant k), \\ \int f(x)\mathrm{d}x(2n \leqslant x \leqslant 2n+1), \qquad \int f(x,y)\mathrm{d}x\mathrm{d}y(x^2+y^2 \leqslant r^2).$$

2.11.4. *Use of the Boolean operations*. The Boolean operations $\vee$ and $\wedge$ often serve (even better than the system given above) to denote '*truncations*' and similar operations. For instance, the function $F(x)$ = '*x provided it is not less than zero or greater than one*' could be written in either of the two ways

$$F(x) = x(0 \leqslant x \leqslant 1) + (x > 1) = 0 \vee x \wedge 1,$$

and the second is clearly simpler. In general, the function which $= f(x)$ but is never less than $m$ or greater than $M$ can be written as $m \vee f(x) \wedge M$, and similarly $m(x) \vee f(x) \wedge$

$M(x)$, so long as we always have $m(x) < M(x)$ (otherwise we would not have $(m \vee f) \wedge M = m \vee (f \wedge M)$, and the notation would not be admissible).

This notation, in our context, will serve in particular for random quantities: we present here a few examples in both notations (and the Boolean form seems to be simpler):

$$X(X \geqslant 0) = 0 \vee X, \qquad X(X \leqslant 0) = 0 \wedge X,$$
$$X = X(X \geqslant 0) + X(X \leqslant 0) = 0 \vee X + 0 \wedge X,$$
$$|X| = X(X \geqslant 0) - X(X \leqslant 0) = 0 \vee X - 0 \wedge X;$$
$$\left.\begin{aligned} &X(0 \leqslant X \leqslant K) + K(X > K) = 0 \vee X \wedge K, \\ &X(|X| \leqslant K) + K\big[(X > K) - (X < -K)\big] = -K \vee X \wedge K, \end{aligned}\right\} (K > 0)$$

and so on.