3

Prevision and Probability

3.1 From Uncertainty to Prevision

3.1.1. So far, even in the way we presented the preceding examples, we have limited ourselves to depicting and representing the situation facing You, when You are interested in distinguishing among a more or less extensive class of alternatives (all those which, in the present state of your information, appear possible to You). This preliminary topic, which we will have to consider more deeply in what follows, is still within the ambit of ordinary logic, the logic of certainty. One should always be careful to distinguish clearly between those things belonging to this domain and those belonging to the *probabilistic* domain – the ambit of the *logic of uncertainty*, the logic of *prevision* – to which we must now turn our attention. It was precisely in order to pin-point this distinction that we decided upon this form of exposition, presenting concepts and related examples which reveal the situation as it is, while leaving undetermined all questions concerning the possible introduction of probability, its conceptual basis and its evaluation. It would certainly be easier, and seemingly more instructive, to go right ahead and take the two steps together, instead of just the one. In other words, we could present right away, fused together in the examples and definitions, both the probability (which answers the need) and the uncertainty (from which the need arises), without first making such a need 'felt', and then pausing to reflect upon it. It is precisely this latter course, however, which must be recommended.

The situation is this: having distinguished the possible cases, and having represented them in the way which seems to You most effective (or in any way convenient to You), if You then wish to restrict yourself to the logic of certainty You have to stop, and consider the question closed. Is this what You *want* to do? And *can* You do it?

For each one of us, it is often the case that we *do not content ourselves* (or are not *able* to content ourselves) with this, and therefore we proceed further. And, strictly speaking, to proceed further means to enter into what we have called the logic of *prevision* (in a sense that we will make clear in order to draw attention to the distinction between this and other interpretations, whose drawbacks must be pointed out).

3.1.2. *Prevision, not prediction*. In order to use this word, 'prevision', it will be necessary to give an absolutely precise meaning to it (and to derived words) and to insist on this meaning and keep it in mind, consistently and scrupulously, in the sequel. It must be distinguished, and in fact contraposed, to another word, which, in everyday language,

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

is perhaps more commonly attributed to it, and for which we will reserve the alternative name, '*prediction*'.

To make a *prediction* would mean (using the term in the sense we propose) to venture to try to 'guess', among the possible alternatives, the one that will occur. This is an attempt often made, not only by would-be magicians and prophets, but also by experts and such like who are inclined to precast the future in the forge of their fantasies.[1] To make a 'prediction', therefore, would not entail leaving the domain of the logic of certainty, but simply including the statements and data which we assume ourselves capable of guessing, along with the ascertained truths and the collected data. It is not enough to tone down the 'prophetic' character of such pronouncements by taking precautions with feelings ('I think', 'perhaps', etc.) as we have already mentioned: either these artificial additions remain without any authentic meaning, or they need to be actually translated into probabilistic terms, substituting prevision in place of prediction.

If we remain within the logic of certainty, such additions not only have no authentic meaning in themselves, but, in point of fact, they render meaningless the entire discussion. If the discussion affirms that something is *true*, and the 'perhaps' means that instead of being true it could also be *false*, this is equivalent to retracting the preceding statement, declaring it to be invalid and unfounded (cancelling it, disowning it). If not, then 'perhaps' should be erased as it might give a false impression of such a retraction.

Alternatively – and this is the approach indicated below, and which corresponds to the subjectivistic conception of probability – the 'perhaps' can be explained as an indication, even if crudely qualitative, of a degree of subjective probability which, if we wished, could be made more precise, and even quantified.

All this would be very clear if there did not exist, unfortunately, in the very field of probability and statistics, certain tendencies to avoid the choice, playing precisely on that ambiguity which we drew attention to earlier, and making it worse. In fact, the ambiguity of the 'perhaps' (which could be innocent, due to simple unwariness) is fraudulently concealed beneath a showy exterior. It is translated into technical terms like 'accept' and 'reject', which neither mean YES or NO with certainty, nor are to be interpreted in a probabilistic sense, but simply lay claim to be themselves 'accepted', rather than 'rejected', without giving to those terms any 'acceptable' meaning whatsoever.

3.1.3. *Prevision*, in the sense in which we have said we want to use this word, does not involve guessing anything. It does not assert – as prediction does – something that might turn out to be true or false, by transforming (over-optimistically) the uncertainty into a

1 Everyone will no doubt have noticed, and had occasion to notice, how often the 'foresights of experts' turn out to be completely different from the facts, sometimes spectacularly so. In the main, this is precisely because they are intended as predictions which 'deduce', more or less logically, a long chain of consequences – still considered necessarily plausible – from the assumed plausibility of an initial hypothesis. Interesting examples of the lack of connection between prevision and reality (in the political field) are pointed out and discussed by B. de Jouvenel in 'Futuribles', *Bulletin Sedeis*, 20 January 1962.

Here also one might note the irrelevant distinction, as far as prevision and prediction are concerned, between the future and the past: the hypothetical reconstructions of murders or historical facts made by detectives, scholars or novelists, based on scanty data and meriting varying degrees of respect, are, in the above sense, 'predictions'.

It is useful to ask oneself, incidentally, whether such 'facile fantasies' are really 'rich fantasies', or rather 'poor fantasies', in that ineptitude or laziness prevents us from seeing how many other possibilities there are, besides the first one we happened to think of.

claimed, but worthless, certainty. It acknowledges (as should be obvious) that what is uncertain is uncertain: in so far as statements are concerned, all that can be said beyond what is said by the logic of certainty is illegitimate. If we think that something might be added, if we think, as we remarked above, that we can *proceed further*, it will necessarily be a question of entering into a completely new field and scheme of things, one which goes beyond the logic of certainty, even if it must be linked to it and superimposed upon it.

When we cease to content ourselves with the logic of certainty, in what sense do we go beyond it? In what sense do You go beyond it? Let us ask ourselves this question. Ask yourself. The thing we are not content with, and neither are You, is the agnostic and undifferentiated attitude towards all those things which, not being known to us with certainty, are uncertain, are possible. There are no degrees[2] of possibility: it is possible (equall possible) that it snows on a winter or summer day; that a great champion or a novice wins the competition; that every student, whether well-prepared or not, will pass an examination; that next Christmas You will find yourself at any place in the world. However, You do not content yourself with this, and, in fact, it is not your real attitude. Faced with uncertainty, one feels, and You feel too, a more or less strong propensity to expect that certain alternatives rather than others will turn out to be true; to think that the answer to a certain question is YES rather than NO; to estimate that the unknown value of a certain quantity is small rather than large.

These attitudes, of ours and of yours, do not lead us – as in the case of someone who claims to make a spot-on prediction – to assert as certain or impossible something which, on the basis of the logic of certainty, is possible but uncertain, and which remains such whatever further assertions or thoughts might be added. Uncertain things remain uncertain, but we attribute to the various uncertain events a greater or lesser degree of that new factor which is extralogical, subjective and personal (mine, yours, his, anybody's), and which expresses these attitudes. In everyday language this is called *probability*, a concept that we shall have to clarify and study. *Prevision*, in the sense we give to the term and approve of (judging it to be something serious, well-founded and necessary, in contrast to prediction), consists in considering, after careful reflection, all the possible alternatives, in order to distribute among them, in the way which will appear most appropriate, one's own expectations, one's own sensations of probability.

We all of us enter into this ambit of prevision in a spontaneous fashion; sometimes without a specific need, for the sole reason that one is interested in the object of uncertainty, that there are desires or hopes that certain alternatives occur, anxieties and fears regarding the occurrence of unfavourable alternatives, and that the weighing up of such hopes and fears matters to one. Sometimes, on the other hand, all one's behaviour may necessarily depend on a comparative evaluation, albeit crude and perhaps unconscious, of the various impending risks, and of the various targets that one can set oneself. In this sense, and because of the enormous range of possibilities, one may find oneself *compelled* to weigh up such evaluations, and to express the prevision. In the case of more important and conscious decisions, one might try to *reason about* each choice, and *weigh up the pros and cons* by means of some criterion or other.

2 In a certain sense, however, there exists a partial order since one could call, 'not less possible' than another, an event which is a consequence of it (in the same way as one could call, 'not less extended' than another one, a set containing it). In both cases, however, no step forward is made towards a comparability or measurability of the 'possibilities' or of the 'extensions'.

3.1.4. *Coherence.* It is precisely in investigating the connection that must hold between evaluations of probability and decision making under conditions of uncertainty that one can arrive at criteria for measuring probabilities, for establishing the conditions which they must satisfy, and for understanding the way in which one can, and indeed one must, '*reason about them*'. It turns out, in fact, that there exist simple (and, in the last analysis, obvious) conditions, which we term conditions *of coherence*: any transgression of these results in decisions whose consequences are manifestly undesirable (leading to certain loss).

The 'one must' is to be understood as 'one must if one wishes to avoid these particular objective consequences'. It is not to be taken as an obligation that someone means to impose from the outside, nor as an assertion that our evaluations are always automatically coherent. On the contrary, it is precisely because this is an area where it is particularly easy to slip into incoherence that it is important to learn the *art of prevision* (to adapt the phrase *Ars Conjectandi*, used by James Bernoulli as the title of the first treatise on the calculus of probability).

Given any set of events whatsoever, the conditions of coherence impose no limits on the probabilities that an individual may assign, except that they must not be in contradiction amongst themselves. Without further delay, we will proceed to the construction of the theory of probability, using as a basis the theory of decision making. For the time being, this will be done in an extremely simplified form, as a preliminary clarification of ideas. In the next paragraph we will discuss certain other aspects of the problem, and then turn to the constructive formulation.

Within this framework, we obtain the greatest insight by considering as a starting point the case of random quantities (especially when we interpret them as random gains). With a more rigorous approach, inspired by decision-theoretic considerations, it is essentially a question of returning to that problem of a *fair evaluation*, or estimation, which, in connection with similar problems of an economic nature, seems to have foreshadowed by centuries the beginnings of the calculus of probability. In this sense, the modern setting of the problem, within decision theory, constitutes, to some extent, a return to its origins.

The definition of the probability of an event will turn out to be contained automatically in that given in the case of random quantities: we simply define events as particular random quantities. From a mathematical viewpoint also, this would appear the appropriate thing to do. In Chapter 2 we saw that in the case of events the most useful arguments, which are very simple if one considers the events as special points in the space of random quantities, are not available if one thinks in terms of the set of events without reference to the space in which this set *is* 'naturally' embedded, and in which it is necessary to *see it* embedded.

Let us proceed then to the matter in hand, starting with the consideration of a *random gain* X: by this we mean a random quantity X having the meaning of gain (the latter intended, of course, in an algebraic sense; a loss is a negative gain). The possible values of X could, therefore, also be negative, either in part, or entirely. We might ask an individual, for example You, to specify the *certain gain* that is considered *equivalent* to X. This we might call the *price* (for You) of X (we denote it by $\mathbf{P}(X)$[3]) in the sense that, on

3 We could write $\mathbf{P}_i(X)$ to emphasize that we are dealing with the evaluation of a particular individual i. This is an unnecessary precaution, however, since it is understood that we are always referring ourselves to the evaluation of a given individual (real or fictitious).

your scale of preference, the random gain X is, or is not, preferred to a certain gain x according to whether x is less than or greater than $\mathbf{P}(X)$. For every individual, in any given situation, the possibility of inserting the degree of preferability of a random gain into the scale of the certain gains is obviously a prerequisite condition of all decision-making criteria. Among the decisions that lead to different random gains, the choice must be the one that leads to the random gain with the highest price. Moreover, this is not a question of a condition but simply of a definition, since the price is defined only in terms of the very preference that it means to measure, and which must manifest itself in one way or another.

In general, it is not true that if one is prepared to buy an article A at the price $\mathbf{P}(A)$ and an article $\boldsymbol{B}$ at the price $\mathbf{P}(B)$, one must be prepared to buy both of them together at a price $\mathbf{P}(A) + \mathbf{P}(B)$. It may happen that the purchase of one of them affects, in various ways, the desirability of the other. Similar qualifications hold if instead of two articles A and B we consider two random gains X and Y; this case will be examined in the next paragraph. In both cases, however, additivity is something more than just an interesting simplifying hypothesis, which may be approximately valid. As we shall see later, provided we modify slightly the way in which the notion of price $\mathbf{P}(X)$ is introduced, additivity will turn out to be an exact property, the foundation of the whole treatment.

3.1.5. *Properties of* $\mathbf{P}$. If You are indifferent to the exchange of X for $\mathbf{P}(X)$ and of Y for $\mathbf{P}(Y)$, then, if we assume the simplifying hypothesis given above, You are also indifferent to the exchange of $X + Y$ for $\mathbf{P}(X) + \mathbf{P}(Y)$. The value for which You are indifferent to the exchange of $X + Y$ is, however, by definition, $\mathbf{P}(X + Y)$; we therefore conclude that

a) *the price* $\mathbf{P}$ *is an additive function*:

$$\mathbf{P}(X+Y)=\mathbf{P}(X)+\mathbf{P}(Y). \tag{3.1}$$

A second property, obvious, but equally fundamental, can be derived by noting that $\mathbf{P}(X)$ must not be less than the lower bound of the set of possible values for X, inf X, nor greater than the upper bound, sup X (otherwise the choice would allow a certain loss). Therefore,

b) *the price* $\mathbf{P}$ *must satisfy the inequality:*

$$\inf X \le \mathbf{P}(X) \le \sup X; \tag{3.2}$$

obviously, this condition only imposes a restriction if the random quantity X is *bounded* in at least one direction (either $\inf X > -\infty$ or $\sup X < +\infty$). Generally, but not always, we will restrict our attention to the bounded case (i.e. bounded from above and below).

When we come to formulate and examine this set-up in a more exhaustive fashion, we shall see that the two extremely simple conditions, (*a*) and (*b*), are *not only necessary but also sufficient* for coherence – that is for avoiding undesirable decisions. This is all that is needed for the foundation of the whole theory of probability: in fact, the definition of probability immediately reduces, as a special case, to that of a price $\mathbf{P}$.

We observe, from (*a*) and (*b*), that the price $\mathbf{P}$ must also be a *linear* function, in the sense that for every real a we have

$$\mathbf{P}(aX)=a\mathbf{P}(X),^4 \tag{3.3}$$

and therefore, more generally,

$$\mathbf{P}(aX+bY+cZ+\ldots)=a\mathbf{P}(X)+b\mathbf{P}(Y)+c\mathbf{P}(Z)+\ldots \tag{3.4}$$

for any *finite* number of summands.

Given this property, it is possible to extend the definition of $\mathbf{P}(X)$ to the case in which X is a random quantity (pure number), or a random magnitude not having the meaning of gain (for instance, time, length, etc.). In fact, it suffices to choose a coefficient a whose dimension is such that aX is a monetary value: for instance, in the cases of time and length we could take Lire/s and \$/cm. We now define $\mathbf{P}(X) = (1/a)\mathbf{P}(aX)$: this is well defined, since the expression is invariant with respect to the choice of a (we can substitute λa in place of a, where λ is a nonzero real number).

In the general case (where we do not have a monetary value), the term 'price' is no longer appropriate: we speak instead of the '*prevision of X*', valid in all cases,[5] and, in particular, of the '*probability of E*' when $X = E$ is an event.

The probability $\mathbf{P}(E)$ that You attribute to an event E is, therefore, the certain gain p that You judge equivalent to a unit gain conditional on the occurrence of E: in order to express it in a dimensionally correct way, it is preferable to take pS equivalent to S conditional on E, where S is any amount whatsoever, one Lira or one million, \$20 or £75. Since the possible values for a possible event E satisfy $\inf E = 0$ and $\sup E = 1$, for such an event we have $0 \leq \mathbf{P}(E) \leq 1$, while necessarily $\mathbf{P}(E) = 0$ for the impossible event, and $\mathbf{P}(E) = 1$ for the certain event.[6]

3.2 Digressions on Decisions and Utilities

3.2.1. In Section 3.1, we have introduced the notions of *prevision* and *probability* by following the path laid down by certain decision-theoretic criteria of an essentially economic nature: the presentation was, however, in a simplified form.

It follows, therefore, that before going any further we should make some comments and give some further details about the theory of decision making, and above all about *utility*. The latter, together with probability, is one of the two notions on which the correct criterion of decision making depends. We warn the reader, however, that this is in

---

4 This is obvious if a is rational, and the extension to every a is straightforward if X is always positive (because then if a lies between a' and a'', we also have aX between $a'X$ and $a''X$). But we can always write $X = Y - Z$, where $Y = X\,(X \geq 0)$ and $Z = -X(X \leq 0)$, and these numbers are always non-negative: $Y = X$ if $X > 0$ and zero otherwise, $Z = -X$ if $X < 0$ and zero otherwise. The conclusion is therefore valid for Y and for Z, and hence for $X = Y - Z$.

5 This corresponds to 'mathematical expectation' in classical terminology, and to 'mean value' in more up-to-date usage. We prefer to reserve the term 'mean value' for *objective* distributions (e.g. statistical distributions).

6 These are the only cases in which the evaluation of the probability is predetermined, rather than permitting the choice of any value in the interval from 0 to 1 (*end-points included*). The predetermination that one meets in these cases arises because there exists no uncertainty and the use of the term probability is redundant. The same thing holds for prevision: $\mathbf{P}(X)$ necessarily has a given value x if and only if X has x as a unique possible value; i.e. if X is not really random. The above is the special case where either $x = 0$ or $x = 1$.

the nature of a digression and anyone not interested in the topic can skip it without any great loss: the details (of a noneconomic nature) that are given in Section 3.3, and in subsequent sections, will prove quite sufficient.

3.2.2. *Operational definitions*. In order to give an effective meaning to a notion – and not merely an appearance of such in a metaphysical–verbalistic sense – an operational definition is required. By this we mean a definition based on a criterion that allows us to measure it.[7] We will, therefore, be concerned with giving an operational definition to the prevision of a random quantity, and hence to the probability of an event.

The criterion, the operative part of the definition which enables us to measure it, consists in this case of testing, through the *decisions* of an individual (which are observable), his *opinions* (previsions, probabilities), which are not directly observable.

Every measurement procedure and device should be used with caution, and its results carefully scrutinized. This is true in physics, despite the degree of perfection attainable, and even more so in a field as delicate as ours, where similar and much more profound difficulties are encountered.

In the first place, if, as is implicit in what we have said so far, we identify, generically, decisions and preferences, then we are ignoring many of the extraneous factors that play a part in decision making. Nobody accepts all the opportunities or bets that he judges favourable, and perhaps we all sometimes enter into situations that we judge unfavourable. To reduce the influence of such factors it is convenient to effect the observations on the phenomena isolated in their most simple forms: this is in fact what we attempt to do when we construct measuring devices. For the purpose of a formal treatment of the topic, we will present (in the next section) two different procedures by means of which we try to force the individual to make conscious choices, releasing him from inertia, preserving him from whim. Of course, we have to establish that the two procedures are equivalent, and this we shall do.

A doubt might remain, however. Are the conclusions that we draw after observing the actual behaviour of an individual, directly making decisions in which he has a real interest, more reliable than those based on the preferences which he expresses when confronted with a hypothetical situation or decision? Both the direct interest and the lack of it might on the one hand favour, and on the other obstruct, the calmness and accuracy, and hence the reliability, of the evaluations. In any case, it is not really a mathematical question: it is useful to be aware of the problem, but it is mainly up to the psychologists to delve further into the matter. We merely note that between the two extreme hypotheses one could consider an intermediate one that might be of interest; the case of an individual being consulted about a decision in which others are interested. This might well lead to responsibility in the judgment without affecting the calmness of the decision maker. In Chapter 5, 5.5.6, we encounter another example which is similar in spirit to the last one: this is where the accuracy of the evaluation is related to one's self-respect in some competitive situation (with prizes which are materially insignificant, but which are related to the significance of the competition).

At this point, the reader may be wondering on what basis individuals do evaluate their probabilities or previsions: the question is not appropriate, however. Firstly, we must

7 See P. Bridgman, *The Logic of Modern Physics*, Macmillan, New York (1927).

attempt to discover opinions and to establish whether or not they are coherent. Only at the second stage, having acquired the necessary knowledge, could we also apply it to investigate these other aspects, and not until it was very much advanced could this be done in a sufficiently satisfactory way (up to and including the rather complicated justifications for the case of evaluations based on frequencies, a case wrongly considered simple).

3.2.3. *Reservations concerning rigidity*. The main question that we have to face in these 'remarks' is the one already mentioned when we expressed reservations about assuming additivity for the price of a random gain: recall that it is this hypothesis which underlies the definition of prevision, and the special case of probability.

It is well known, and indeed obvious, that usually this is not realistic because of the phenomenon of risk aversion (or occasionally its opposite, but we shall not bother with such cases). In fact, as we already noted in effect when we introduced it, the hypothesis of additivity expresses an assumption of *rigidity* in the face of risk. Let us now try to make this clear. As a preliminary, it will suffice to restrict ourselves to simple examples that are within our present scope. These will be sufficient to show that in order to obtain a formulation which is completely satisfactory from the economic point of view, it is necessary to eliminate such rigidity by introducing the notion of *utility*. On the other hand, they will also show that one is able to manage without this notion, except when occupied with applications of an expressly economic nature.

Suppose that You are faced with two eventualities that You judge equally probable: taking the standard example, it could be a question of Heads or Tails. Given the hypothesis of rigidity in the face of risk, You should be indifferent between 'receiving with certainty a sum S, or twice the sum if a particular one of the two possible cases occurs': likewise, between 'losing with certainty a sum S, or twice the sum if a particular one of the two possible cases occurs'; and similarly between 'accepting or not accepting a bet which, in the two possible cases, would lead either to a loss, or to a gain, of the same sum S.' This much is obvious, but in any case we shall carry out the calculations as an exercise. Let us denote by A and B the two events: $A + B = 1$ because one and only one of the two occurs. Their probabilities, being supposed equal, must each have the value $\frac{1}{2}$, since $\mathbf{P}(A) = \mathbf{P}(B)$, and $\mathbf{P}(A) + \mathbf{P}(B) = \mathbf{P}\,(A + B) = \mathbf{P}(1) = 1$. It follows that cases of so-called indifference simply imply the equality of the following: S and $(2S)/2$, $-S$ and $-(2S)/2$, 0 and $\frac{1}{2}S + \frac{1}{2}(-S)$ (since, for instance, the gain $2S$ conditional on the event A is the random quantity $X = 2SA$, $\mathbf{P}(X) = \mathbf{P}(2SA) = 2S\mathbf{P}(A) = 2S\,.\left(\frac{1}{2}\right)$.

If instead, as is likely, You are risk averse, then in all cases You will prefer the *certain* alternative to the *uncertain* one (the form and extent of the aversion will depend upon your temperament, or perhaps be influenced by your current mood, or by some other circumstance). To arrive at the actual indifference, You would content yourself with receiving with certainty a sum S' (less than S) in exchange for the hypothetical gain $2S$; You would be disposed to pay with certainty a sum S'' (greater than S) in order to avoid the risk of a hypothetical loss $2S$; You would pay a certain penalty K in order to be released from any bet where the gain and loss are, in monetary terms, symmetric.

This means, however, that by virtue of risk aversion one has symmetry in the scale in which one's judgments of indifference are based: that is equal levels in passing from 0 to S' and from S' to $2S$, or in passing from $-2S$ to $-S''$ and from $-S''$ to 0, or in passing from $-S$ to $-K$ and from $-K$ to S. The scale no longer coincides with the monetary one, as in the case of rigidity. In short, as far as we are concerned, things proceed as if successive increments of equal monetary value had for You smaller and smaller subjective value or *utility*. This term – often used in a similar sense, but in a questionable form, in economic science – has been rehabilitated and adopted with the specific meaning derived from the present considerations about risk.

3.2.4. *The scale of utility*. The above considerations enable us to construct a scale of utility; that is a function $U(x)$, the utility of the gain x, whose increments, $U(x_{i+1}) - U(x_i)$, are equal when, and only when, we are indifferent between the corresponding increments of monetary gain, $x_{i+1} - x_i$. We could proceed, for instance, by dividing an interval into two 'indifferent increments', in the way indicated in the examples above, and in the same way obtain subdivisions into 4, 8, 16,…, parts. It would be more appropriate, instead of considering the variable x representing the gain, to take $f + x$, where f is the individual's 'fortune' (in order to avoid splitting hairs, inappropriate in this context, one could think of the value of his estate). Anyway, it would be convenient to choose a less arbitrary origin in order to take into account the possibility that judgments may alter because in the meantime variations have occurred in one's fortune, or risks have been taken, and in order not to preclude for oneself the possibility of taking these things into account, should the need arise. Indeed, as a recognition of the fact that the situation will always involve risks, it would be more appropriate to denote the fortune itself by F (considering it as a random quantity), instead of with f (a definite given value).

What we have said concerning the scale of utility makes it intuitively clear – and this is sufficient for the time being – that if, in order to define 'price', we refer to this scale rather than to the monetary scale, then additivity holds. In fact, one might say that such a scale is by definition the monetary scale deformed in such a way as to compensate for the distortions of the case of rigidity which are caused by risk aversion. The formulation put forward in the preceding paragraph could, therefore, be made watertight, and this we will do shortly by working in terms of the utility instead of with the monetary value. This would undoubtedly be the best course from the theoretical point of view, because one would construct, in an integrated fashion, a theory of decision making (of the criteria of coherent decisions, under conditions of certainty or uncertainty), whose meaning would be unexceptionable from an economic viewpoint, and which would establish simultaneously and in parallel the properties of *probability* and *utility* on which it depends. The fundamental result lies, in fact, in recognizing that *the criteria of coherent decision making are precisely those which consist of the choice of any evaluation of the probabilities and any utility function* (with the necessary properties) *and in fixing as one's goal the maximization of the prevision of the utility*. Of course, it is possible to behave coherently with respect to decisions and preferences without knowing anything about probability and utility. The fact is, however, that in this case one must behave *as if* one is acting in the above manner, *as if* obeying an evaluation of probability and a scale of utility underlying one's way of thinking and acting (even if without realizing it). Provided one could succeed in exploring these activities in an appropriate way, it might be possible to trace back and individuate the two components.

This unified approach to an integrated formulation of decision theory in its two components was put forward by F.P. Ramsey (1926) and rigorously developed by L.J. Savage (1954). However, there are also reasons for preferring the opposite approach, the one which we attempt here. This consists in setting aside, until it is expressly required, the notion of utility, in order to develop in a more manageable way the study of probability.

3.2.5. *An alternative approach.* The idea underlying this alternative approach stems from the observation that the hypothesis of rigidity, as considered above, is acceptable in practice – even if we stick to monetary values – provided the amounts in question are not 'too large'. Of course, the proviso has a relative and approximate meaning: relative to You, to your fortune and temperament (in precise terms, to the degree of convexity of your utility function U); approximate because, in effect, we are substituting in place of the segment of the curve U which is of interest the tangent at the starting point. Clearly, the smaller the range considered, the more satisfactory is the approximation. With this in mind, we might consider replacing the previous definition of $\mathbf{P}(X)$ – which we temporarily distinguish, denoting it by $\mathbf{P}^*(X)$ – with a new one, which we define by means of the relationship:

$$\mathbf{P}(X) = \lim_{a \to 0}\left(\frac{1}{a}\right)\mathbf{P}^*(aX).$$

Instead, we prefer a less orthodox but more natural and manageable solution, which consists of not changing anything, but merely remarking that in economic examples one must remain within appropriate limits (which, as an aid to understanding, we call 'everyday affairs').

There are several reasons behind this choice (and, more generally, behind rejecting the standard method of considering both probability and utility together, right from the very beginning).

Firstly, on a purely formal level, there is an objection to taking the passage to the limit so seriously as to base a definition on it: in fact, if a becomes too small an evaluation loses, in practice, any reliability. This is the same phenomenon that one encounters when attempting to define density, although the underlying reasons are different. One needs to consider the ratios mass/volume for neighbourhoods small enough to avoid macroscopic inhomogeneities, but not so small as to be affected by discontinuities in the structure of matter. We accept that once we are in the area of mathematical idealization we can leave out of consideration adherence to reality in every tiny detail: on the other hand, it seems rather too unrigorous to act in this way when formulating that very definition that should provide the connection with reality.

This does not mean that it is not useful to accept the form of the passage to the limit (as an innocuous and convenient assumption, although not appropriate to fulfil the function of a definition). In any case, let us suppose that we have introduced the linear prevision $\mathbf{P}(X)$, and that we know the utility function U, which, for the sake of simplicity, we now take to be expressed as a function of the gain x. Then the original $\mathbf{P}(X)$ as it actually turns out to be, assuming that the hypothesis of rigidity is not satisfied (this is denoted above by $\mathbf{P}^*$, but from now on we denote it by $\mathbf{P}_U$), can be expressed immediately as a transform of $\mathbf{P}$ by means of U:

$$\mathbf{P}_U(X) = U^{-1}\{\mathbf{P}[U(X)]\}. \tag{3.5}$$

In the standard case, where U is convex, we have $\mathbf{P}_U(X) < \mathbf{P}(X)$, as noted above, and as can be seen from the theory of associative means (Chapter 2, 2.9.3). In order to be able to distinguish between the two concepts, when referring to them, we will say that:

a transaction is depending on whether	$\mathbf{P}_U$	*indifferent,* remains constant,	*advantageous,* or increases,	*disadvantageous* or decreases;
a transaction is depending on whether	$\mathbf{P}$	*fair,* remains constant,	*favourable,* or increases,	*unfavourable* or decreases.

A fair transaction is such for everyone agreeing on the same evaluation of probabilities, even for the other contracting party ($\mathbf{P}(-X) = -\mathbf{P}(X)$); an indifferent transaction is not such as U varies, and cannot be such for both contracting parties if they both have convex utility functions (in this case $\mathbf{P}_U(-X) < \mathbf{P}(-X) = -\mathbf{P}(X) < -\mathbf{P}_U(X)$).

3.2.6. *Some further remarks.* Finally, let us turn to the other reasons for preferring this approach: these are essentially concerned with simplicity. The separation of probability from utility, of that which is independent of risk aversion from that which is not, has first of all the same kind of advantages as result from treating geometry apart from mechanics, and the mechanics of so-called rigid bodies without taking elasticity into account (instead of starting with a unified system).

The main motivation lies in being able to refer in a natural way to combinations of bets, or any other economic transactions, understood in terms of monetary value (which is invariant). If we referred ourselves to the scale of utility, a transaction leading to a gain of amount S if the event E occurs would instead appear as a variety of different transactions, depending on the outcome of other random transactions. These, in fact, cause variations in one's fortune, and therefore in the increment of utility resulting from the possible additional gain S: conversely, suppose that in order to avoid this one tried to consider bets, or economic transactions, expressed, let us say, in 'utiles' (units of utility, definable as the increment between two fixed situations). In this case, it would be practically impossible to proceed with transactions, because the real magnitudes in which they have to be expressed (monetary sums or quantity of goods, etc.) would have to be adjusted to the continuous and complex variations in a unit of measure that nobody would be able to observe.

Essentially, our assumption amounts to accepting as practically valid the hypothesis of rigidity with respect to risk: in other words, the identity of monetary value and utility[8] within the limits of 'everyday affairs'. One should be concerned, however, to check whether this assumption is sufficiently realistic within a wide enough range: actually, it seems safe to say that under the heading of 'everyday affairs' one can consider all those transactions whose outcome has no relevant effect on the fortune of an individual (or firm, etc.), in the sense that it does not give rise to substantial improvements in the situation, nor to losses of a serious nature.

There is no point in prolonging this discussion, but it seems appropriate to mention an analogy from economics, and one from insurance: these – in the same spirit as the

8 Except for (obviously inessential) changes of origin and unit of measurement.

preceding geometrical–mechanical analogy – are sufficient to clarify the question, both from a conceptual and practical point of view. Using the prices $\mathbf{P}(X)$ as they appear in our hypothesis of rigidity is to do the same thing as one does in economics when one considers the total price of a set of goods, of given amounts, on the basis of the unit prices in force at the time, without taking into account the variation that a possible transaction would cause by changing the supply and demand situation. On the other hand, these variations are only noticeable if the quantities under consideration are sufficiently large. Even more apposite is the example of actuarial mathematics: indeed, the latter is nothing other than a special case of the theory we are discussing. In the main, it is traditionally concerned with the terms of an insurance under *fair* conditions ('pure' is the usual terminology: pure premium, etc.), and only in special cases – for instance, the theories dealing with the risk of the insurer, or with the advantage for those exposed to risk in insuring themselves – does one speak in terms of utility (or something equivalent, if such a notion is not introduced explicitly). Notwithstanding the fact that this stems less from deliberate choice than from a tradition that lacks an awareness of the questions involved, the 'rigid' approximation has turned out to be satisfactory for the greater part of this most classical field of application of the calculus of probability to economic questions. We intend to use only the simplified version; the above considerations suggest that this is a reasonable thing to do.[9]

On the other hand, we shall see (in Chapter 5) how, although starting from the hypothesis of rigidity, one can arrive at the evaluation of probabilities by means of criteria which are neutral with respect to it. The method, which takes as its basis the most meaningful concept, and then clarifies it by means of this simplifying hypothesis, therefore achieves its objective without prejudicing the conclusions.

3.3 Basic Definitions and Criteria

3.3.1. We must now translate into actual definitions and proofs those things that we have hitherto put forward in an introductory form, bringing in any necessary refinements, and beginning the developments.

We have given some idea of what a *prevision function* $\mathbf{P}$ is, and what conditions it must satisfy in order to be *coherent*. The function $\mathbf{P}$ represents the opinion of an individual who is faced with a situation of uncertainty. To each random magnitude X, there corresponds the individual's evaluation $\mathbf{P}(X)$, the *prevision* of X, whose meaning, operationally, reduces, in terms of gain, to that of the (fair) *price* of X. This includes, in particular, the special case of *probability* (which is the more specific name given to prevision when X is an event). A prevision function $\mathbf{P}$ is *coherent* if its use cannot lead to an inadmissible decision (i.e. such that a different possible decision would have certainly led to better results, whatever happened). We have remarked already that coherence reduces to linearity and convexity.

3.3.2. In order to fix the formulation in a precise way, we will now put forward two *criteria* (in the sense of *devices* or *instruments* for obtaining a measurement). Each one furnishes an *operational definition* of probability or prevision $\mathbf{P}$, and,

9 For all these topics see de Finetti–Emanuelli (1967), Part I.

together with the corresponding *conditions of coherence*, can be taken as a foundation for the entire theory of probability.

Let us recall that the term 'operational' applies to those definitions that allow us to reduce a concept not merely to sentences, which might have only an apparent meaning, but to actual experiences, which are at least conceptually possible. Think of Einstein's definition of 'simultaneity' by means of signals: until that time no-one had even doubted that the term lacked an absolute meaning. That definitions should be operational is one of the fundamental needs of science, which has to work with notions of ascertained validity, in a pragmatic sense, and which must not run the risk of taking as concepts illusory combinations of words of a metaphysical character.

In our case, for the definition of $\mathbf{P}(X)$, it is a question of stating exactly what 'the rules of the game' are. To state, in other words, what, in the application of a given criterion, are the practical consequences that You know You must accept, and which You do accept, when You enunciate your evaluation of $\mathbf{P}(X)$ (whose meaning as 'price' is already essentially given). From a conceptual point of view, in the case of coherence too the pointers given in Section 3.1.5 are sufficient in themselves. To make them explicit in a compact form for specific criteria provides, however, a more incisive schematization of the theory by reducing it to a really small nucleus of initial assumptions.

3.3.3. As far as the extension of the domain of definition of a function of prevision $\mathbf{P}$ is concerned, we assume that in principle $\mathbf{P}$ could be evaluated (by You, by anybody) for *every* event E or random quantity X: this is in contrast to what is assumed in other theories and so it is appropriate to point it out explicitly. It will be sufficient for You to place yourself under the restriction of a certain criterion, which we shall soon make explicit, and being forced to answer – that is to make a choice among the alternatives at your disposal – to reveal your evaluation of $\mathbf{P}(X)$ (or, in particular, of $\mathbf{P}(E)$). This is valid, as we have said, *in principle*: in other words, we intend not to acknowledge any distinction according to which it would make sense to speak of probability for some events, but not for others.[10] On the other hand, however, we certainly do not pretend that $\mathbf{P}$ could actually be imagined as determined, by any individual, for *all* events (among which those mentioned or thought of during the whole existence of the human race only constitute an infinitesimal fraction, even though an immense number). On the contrary, we can at each moment, and in every case, assume or suppose $\mathbf{P}$ as defined or known for all (and only) the random quantities (or, in particular, events) belonging[11] to some completely arbitrary set $\mathscr{X}$: for instance, those for which we know the evaluation explicitly expressed by the individual under consideration.

Without leaving this set, whatever it may be, we can recognize whether or not $\mathbf{P}$ includes any incoherence; if so, the individual, when made aware of this fact, should eliminate it, modifying his evaluations after reconsideration. The evaluation is then coherent and can be extended to any larger set whatsoever: the extension will be uniquely determined up to the point that the coherence demands and is, to a large extent, more or less arbitrary outside that range. One can only proceed, therefore, by interrogating the individual and alerting him if he violates coherence with respect to the preceding evaluations.

10 For instance, the two following distinctions are quite common: yes for 'repeatable' events, no for 'single' instances; yes if X belongs to a measurable set I, no otherwise. See Appendix.

11 And not, necessarily, belonging to something reducible to a ring (or to a σ-ring) of events. Again, see Appendix.

Among the answers that do not make sense, and cannot be admitted, are the following: 'I do not know', 'I am ignorant of what the probability is', 'in my opinion the probability does not exist'. Probability (or prevision) is not something that in itself can be known or not known: it exists in that it serves to express, in a precise fashion, for each individual, his choice in his given state of ignorance. To imagine a greater degree of ignorance that would justify the refusal to answer would be rather like thinking that in a statistical survey it makes sense to indicate, in addition to those whose sex is unknown, those for whom one does not even know 'whether the sex is unknown or not'.

Other considerations and restrictions may enter in if we consider functions of probability defined other than as an expression of the opinion of a given individual. If, for instance, after having considered and interrogated many individuals, we want to study 'their common opinion', $\mathbf{P}$, this will only exist in the domain of those X for which all the $\mathbf{P}_i(X)$ coincide (in this way defining $\mathbf{P}(X)$), and will not exist elsewhere. We can also confine ourselves (there is nothing to prevent us) to evaluations which conform to more restrictive criteria to which one would prefer to limit the investigation, excluding in this way events for which one would like to say that the probability 'does not exist' or 'is not known', knowing all along that such motivations remain, nonetheless, meaningless within the present formulation. I may please a friend of mine by not inviting along with him a person whom he judges 'a jinx', without myself believing that such things exist, nor understanding how others can believe in it.

As far as *coherence* is concerned, we will again underline here in what sense the notion is and must be *objective*. The conditions of coherence must exclude the possibility of certain consequences whose unacceptability appears expressible and recognizable to everyone, independently of any opinions or judgments they may have regarding greater or lesser 'reasonableness' in the opinions of others. Let this be said in order to make clear that such conditions, although *normative*, are not (as some critics seem to think) unjustified impositions of a criterion that their promoters consider 'reasonable': they merely assert that 'you must avoid this if you do not want …' (and there follows the specification of something which is obviously undesirable). We will see this immediately – note it well! – in the two criteria we are about to put forward.

3.3.4. *Criteria for the evaluations.* We now present the details of the two criteria mentioned above; each will consist of the following:

- a *scheme of decisions* to which an individual (it could be You) can subject himself in order to reveal – in an operational manner – that value which, *by definition*, will be called his *prevision* of X, or in particular his *probability* of E,
- and a *condition of coherence* that enables one to distinguish (so that the distinction has an objective meaning) whether an individual's set of previsions is coherent, and therefore acceptable, or, conversely, intrinsically contradictory.

In both cases, the prevision of X will be a value $\bar{x}$, which can be chosen at will as an 'estimate' of X; along with such a choice goes the necessity of making precise, according to which scheme is used, the otherwise completely indeterminate[12] meaning of the

12 Or, even worse, open to being interpreted as 'prediction'!

word 'estimate'. To anticipate the outcome in words, both criteria start by considering the random magnitude given by the difference, or deviation, $X - \bar{x}$, between the actual value X and that chosen by You. Both lead to the same $\bar{x}$ if applied coherently.

The first criterion stipulates that You must accept a bet proportional to $X - \bar{x}$, in whatever sense chosen by your opponent (i.e. positively proportional either to $X - \bar{x}$ or to $\bar{x} - X$). This means that there is no advantage to You in deviating, one way or the other, from the value that makes the two bets indifferent for You; otherwise, one or other would be unfavourable to You, and the opponent could profit from this by an appropriate choice.[13]

The second criterion stipulates that You will suffer a penalty (positively) proportional to the square of the deviation $(X - \bar{x})^2$, increasing as one deviates in either direction from the actual value.

This is evident if one recalls the properties of the barycentre (stable equilibrium, minimum of the moment of inertia), which give an analogy and, in fact, a perfect interpretation. Those who already know something about probability or statistics will be well acquainted with the fact that these properties characterize $\mathbf{P}(X)$. The latter is usually called '(mathematical) expectation', or 'mean value', and is denoted by $\mathbf{E}(X)$ or $\mathbf{M}(X)$: the only novelty lies in making use of it as an operational and direct definition of $\mathbf{P}(X)$, and in particular of probability. Given the probabilities of all possible values (if they are finite in number), it is clear how $\mathbf{P}(X)$ can be expressed as a function of them: the extension of this result to the general case will be immediate when we introduce the notion of a 'probability distribution' (see Chapter 6). In the latter approach, however, one introduces the simpler notion (that of 'prevision') by means of the more complicated one (that of 'distribution'), which itself becomes a prerequisite, and forces us to use more advanced mathematical tools (Stieltjes integrals) than necessary. The same thing happens in the case of a solid body: the barycentre is easily determined and, it might be said, is always useful; the exact distribution of mass can never be determined in practice, and is of relatively little interest.

Two further remarks. Firstly, let us recall 'the hypothesis of rigidity with respect to risk', which we continue to assume in what follows (not without noting where appropriate, under the heading of 'Remarks', any implications of this hypothesis at those points where it merits attention). In order to fulfil more easily the resulting requirement of considering only 'moderate amounts', and to omit certain delicate points which are better reconsidered later on (in Section 3.11, etc.), we restrict ourselves for the time being to *bounded* random magnitudes (i.e. those whose possible values are contained in some interval; in other words, $-\infty < \inf X, \sup X < +\infty$).

Concerning preferences for one or other of the two criteria of definition, it is merely a question of individual taste, since – as we have stated, and will later show – the two definitions (together with their respective conditions of coherence) are equivalent. The first has a meaning that is slightly more immediately intuitive, but, as far as actual deductions are concerned, the second is more meaningful and fits better into a decision-theoretic framework. A third criterion, which has useful applications, will be derived in Chapter 5, but does not lend itself to an autonomous presentation.

13 This is the same criterion as 'divide the cake into two parts and I will choose the larger', which ensures that the person dividing it does so into parts he judges to be equal.

3.3.5. *The first criterion*. Given a random quantity (or random magnitude) X, You are obliged to choose a value $\bar{x}$, on the understanding that, after making this choice, You are committed to accepting any bet whatsoever with gain $c(X - \bar{x})$, where c is arbitrary (positive or negative) and at the choice of an opponent.

Definition. $\mathbf{P}(X)$, the prevision of X according to your opinion, is by definition the value $\bar{x}$ that You would choose for this purpose.

Coherence. It is assumed that You do not wish to lay down bets which will with *certainty* result in a loss for You.[14] A set of your previsions is, therefore, said to be *coherent* if among the combinations of bets which You have committed yourself to accepting there are none for which the gains are *all uniformly negative*.[15]

Analytic conditions. Expressed mathematically, this means that we must choose the values $\bar{x}_i = P(X_i)$ such that there is no linear combination

$$Y = c_1(X_1 - \bar{x}_1) + c_2(X_2 - \bar{x}_2) + \ldots + c_n(X_n - \bar{x}_n)$$

with sup Y negative (conversely, inf Y cannot be positive, because then sup $(-Y) = -$inf Y would be negative).

Remark. Observe the objective character of these conditions, revealed by the fact that only 'possible values' are referred to.

3.3.6. *The second criterion*. You suffer a penalty L[16] proportional to the square of the difference (or deviation) between X and a value $\bar{x}$, which You are free to choose for this purpose as you please:

$$L = \left(\frac{X - \bar{x}}{k}\right)^2$$

(where k, arbitrary, is fixed in advance, possibly differing from case to case).[17]

Definition. $\mathbf{P}(X)$, the prevision of X according to your opinion, is the value $\bar{x}$ which You would choose for this purpose.

Coherence. It is assumed that You do not have a preference for a given penalty if You have the option of another one which is *certainly* smaller. Your set of previsions is therefore said to be *coherent* if there is no other possible choice which would certainly lead to a uniform reduction in your penalty.

Analytic conditions. The definition of coherence implies that there exist no values x_i^* which, when substituted for the chosen $\bar{x}_i = \mathbf{P}(X_i)$, lead to the penalty

14 Giving rise to what is sometimes called a 'Dutch Book'.
15 The reason why we cannot simply say 'all negative' (i.e. <0), but must add 'uniformly' (i.e. $< -\varepsilon$ with ε positive) will be given later (for the time being we do not worry about the finer points). By 'combinations' we always mean linear combinations *of a finite number* of the bets even if there are infinitely many of them).
16 From *Loss*, the terminology introduced by A. Wald.
17 It is convenient to think of k as being homogeneous with X, so that the expression turns out to be a pure number; with the further understanding that we multiply by a monetary unit u, L has the dimension of a monetary value. This avoids the complication of writing it, or assuming it as included in k, by conjuring up a strange factor of dimension $u^{1/2}$.

$$L^* = \sum_i \left(\frac{X_i - x_i^*}{k_i} \right)^2$$

being uniformly less than

$$L = \sum_i \left(\frac{X_i - \bar{x}_i}{k_i} \right)^2,$$

for *any possible points* (X_i, X_2, ..., X_n); that is belonging to the set $\mathfrak{Q}$.

Remark. As for the first criterion.

3.3.7. *The equivalence of the two criteria.* The identity of the previsions given by the two criteria can be verified immediately.

Let $\bar{x}$ be the prevision of X based on the first criterion, and $\bar{\bar{x}}$ that based on the second; this implies, respectively, that:

i) in the first case, the random gain X is judged equivalent to the certain gain $\bar{x}$ (hence: preferable to each $x < \bar{x}$, but not to any $x > \bar{x}$);
ii) in the second case, the gain $-(X - \bar{\bar{x}})^2$ – negative, since a penalty ! – is judged preferable to any other $-(X - x)^2$ with $x \neq \bar{\bar{x}}$; in other words, the gain

$$G = \left(X - x\right)^2 - \left(X - \bar{\bar{x}}\right)^2$$

is preferred to 0 (for all $x \neq \bar{\bar{x}}$).

More generally, let us compare preferences between the penalties corresponding to any two values of x, say $x = a$ and $x = b$, and let us denote by $c = \frac{1}{2}(a + b)$ the mid-point of the interval $[a, b]$.

The choice of a is preferred to that of b, if the gain $G = (X - b)^2 - (X - a)^2$ if preferred to 0; in other words, expanding, if

$$\begin{aligned} G &= \left(X^2 - 2bX + b^2\right) - \left(X^2 - 2aX + a^2\right) = 2\left(a - b\right)X - \left(a^2 - b^2\right) \\ &= 2\left(a - b\right)\left(X - c\right) \end{aligned}$$

is preferred to 0. Preferring G to 0 means that $\mathbf{P}(G) > 0$; on the basis of the first criterion it turns out that $\mathbf{P}(G) = 2(a - b)(\bar{x} - c)$, an expression which is positive if $a > b$ and $\bar{x} > c$, or, conversely, if $a < b$ and $\bar{x} < c$. In other words, in either case, if $\bar{x}$ lies in the subinterval between c and a; that is if $\bar{x}$ is *closer* to a than it is to b.

Our assertion is an obvious corollary of this result (which it seemed useful to put forward in this more general form): the optimal choice, $x = \bar{\bar{x}}$, is given by $\bar{\bar{x}} = \bar{x}$.

The equivalence of the conditions of coherence can also be verified by expansions of this sort (and we shall do so, writing them out in full, for those who would like to check them and apply them directly). Conceptually, however, we can make everything incomparably easier, and intuitively meaningful, by presenting an obvious geometrical interpretation.

3.4 A Geometric Interpretation: The Set $\mathscr{P}$ of Coherent Previsions

3.4.1. Any prevision in the linear ambit $\mathscr{A}$ of the n random quantities $X_1, X_2,\ldots, X_n$ consists in fixing, in the n-dimensional space with coordinates $x_1, x_2,\ldots, x_n$ (the linear ambit $\mathscr{A}$), the n values $\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_n$, where $\bar{x}_i = \mathbf{P}(X_i)$, and hence corresponds to a point in the said space. The conditions of coherence state – as we shall immediately verify – that *the set $\mathscr{P}$ of coherent previsions is the closed convex hull of the set $\mathscr{Q}$ of possible points.*

For the first criterion: in a form that is more directly suited to the purpose in hand, the necessary and sufficient condition for coherence can be expressed by saying that *every linear relation (or inequality) between the random quantities X_i*

$$c_1X_1 + c_2X_2 + \ldots + c_nX_n = c \;\left(\text{or } \geqslant c\right)$$

must be satisfied by the corresponding previsions $\mathbf{P}(X_i)$:

$$c_1\mathbf{P}(X_1) + c_2\mathbf{P}(X_2) + \ldots + c_n\mathbf{P}(X_n) = c \;\left(\text{or } \geqslant c\right).$$

Geometrically, a point P represents a coherent prevision if and only if *there exists no hyperplane separating it from the set $\mathscr{Q}$ of possible points*; this characterizes the points of the convex hull.

For the second criterion: here one introduces into the (*affine*!) linear ambit $\mathscr{A}$ a *metric* of the form $\rho^2 = \sum_i (x_i/k_i)^2$, setting

$$\begin{aligned}\text{'penalty'} = L &= (P-Q)^2 \\ &= \text{'the square of the distance between the prevision} \\ &\quad \text{-point } P \text{ and the outcome-point } Q\text{, according to the given metric'.}\end{aligned}$$

The necessary and sufficient condition for coherence requires, in geometrical terms, that *P cannot be moved in such a way as to reduce the distance from all points* Q; this is another characterization of the convex hull.[18] Further explanations, and diagrams in the simple cases, are given in Chapter 5, Section 5.4.

3.4.2. *Other interpretations.* Every prevision-point P, which is admissible in terms of coherence, is a barycentre of possible points Q_j with suitable weights (or is a

18 If we move the point P to another position P^*, its distance from a generic point A increases or decreases depending on whether A is on the same side as P or P^*, with respect to the hyperplane that bisects the segment PP^* orthogonally.

If P is not in the convex hull of $\mathscr{Q}$ there exists a hyperplane separating it from $\mathscr{Q}$. Moving P to P^*, its orthogonal projection onto such a hyperplane, diminishes its distance from all points Q in $\mathfrak{Q}$ (which are on the opposite side). More precisely, the diminution of the penalty, $L - L^*$, i.e. the square of the distance, is always at least $(P - P^*)^2$: in fact, $(Q - P)^2 - (Q - P^*)^2 = [(Q - P^*) - (P - P^*)]^2 - (Q - P^*)^2 = (P - P^*)^2 - 2(Q - P^*) \times (P - P^*)$, and this scalar product is negative, since the component of the first vector parallel to the second is in the opposite direction.

Suppose instead that P belongs to the convex hull of $\mathscr{Q}$. Then to whatever point P^* we move P, it always follows that for some point Q the distance increases: if, with respect to the bisecting hyperplane, they were all on the same side as P^*, the point P would be separated from the convex hull of $\mathscr{Q}$, contrary to the hypothesis.

limit-case[19]). On the other hand, the possible points are themselves particular cases of previsions; degenerate cases, in that the probability is concentrated at a unique point Q_j. In words, one could say, according to this interpretation, that *a prevision turns out to be a mixture of possibilities*.

Of course, one can also form linear combinations of different coherent previsions (with non-negative weights, summing to 1) again obtaining coherent previsions. More generally, if $\mathscr{P}_0$ is any set of coherent previsions, then its closed convex hull is also a set of coherent previsions, the *mixtures* of those in $\mathscr{P}_0$. Let us denote it by $\mathscr{P}_1$.

3.5 Extensions of Notation

It is convenient, in addition to being natural (and also useful for compactness of notation), to exploit the linear structure of **P** in order to extend the range of applicability of this symbol to any random elements whatsoever belonging to a linear space (vectors, matrices, n-tuples of numbers or magnitudes, functions, etc.), or even just to a linear manifold (a linear subspace which also contains the zero: for example, the points of a space in which the differences between points $\mathbf{u} = A - B$, constitute a linear space of vectors).

As a formal definition, it is sufficient to state that **P** is always intended to be *linear*, so that if f is any linear function – that is $f(A)$ is a scalar linear function of the points or elements A of our linear space or manifold – we have $f(\mathbf{P}(A)) = \mathbf{P}(f(A))$. For practical purposes, it is enough to note that **P** operates on the components or coordinates, so that, if

$$A = 0 + X\mathbf{i} + Y\mathbf{j} + Z\mathbf{k}\big(or, \text{ in conventional notation, } A = (X, Y, Z)\big),$$

we could write

$$\mathbf{P}(A) = \mathbf{P}(X)\mathbf{i} + \mathbf{P}(Y)\mathbf{j} + \mathbf{P}(Z)\mathbf{k} = \bar{x}\mathbf{i} + \bar{y}\mathbf{j} + \bar{z}\,\mathbf{k}$$

(in other words, $\mathbf{P}(X, Y, Z) = (\mathbf{P}(X), \mathbf{P}(Y), \mathbf{P}(Z))$).

A case of particular importance is the following: if Z is a *complex* random quantity, and we denote by X and Y, respectively, the real and imaginary components, its prevision will be

$$\mathbf{P}(Z) = \mathbf{P}(X + \mathrm{i}Y) = \mathbf{P}(X) + \mathrm{i}\mathbf{P}(Y) = \bar{x} + \mathrm{i}\bar{y}.$$

As a practical rule, it is sufficient to replace the random component X by the corresponding prevision $\bar{x}$; for example:

$$\mathbf{P}(X_1, X_2, \ldots, X_n) = (\bar{x}_1, \bar{x}_2 \ldots, \bar{x}_n), \qquad \mathbf{P}\big(\|X_{rs}\|\big) = \|\bar{x}_{rs}\|, \qquad \text{etc.}$$

19 To be precise: either they can be obtained as barycentres of most $n + 1$ points Q_j (in the n-dimensional space), or they are *adherent* points of $\mathscr{Q}$ (but not belonging to $\mathscr{Q}$). For instance, we could think of $\mathscr{Q}$ as the set of points on the circumference of a circle, having rational angular distance from one of its points (with respect to the complete angle). We are in the plane, $n = 2$, and each point inside the circle is inside triangles with vertices in $\mathscr{Q}$: hence it is the barycentre of $3 = n + 1$ points (two would suffice if it were on chords connecting rational points, and only one if it coincided with such a point). The points, which are on the circumference, but not rational, are required in order to complete the closed convex hull: they are adherent points of $\mathscr{Q}$ (i.e. there are points of $\mathscr{Q}$ in each of their neighbourhoods).

In the case of a random function $X(t)$ (where t is the independent variable; for example, time) it would appear to be an unnecessary subtlety (but it is not) to say that one could write $f = \mathbf{P}(X)$ to mean that f is the function which for each t gives $f(t) = \mathbf{P}(X(t))$. It would be a little more explicit to write $f(\cdot) = \mathbf{P}(X(\cdot))$ in order to indicate that it is a question of operating on a variable whose position is denoted by the point. Here, however (at least if one does not want to be limited to considering not more than a finite number of t_h at once), one would step outside of the ambit in which, for the time being, we have expressed our intention of remaining.

3.6 Remarks and Examples

3.6.1. The properties that we have established in Section 3.4 could be said to contain the whole calculus of probability, even though we have not as yet mentioned probability, except to point out that it is a special case of prevision. Sections 3.8 and 3.9 will be devoted to this special case, giving it the attention it merits. However, from our point of view it turns out to be better formulated and much clearer if embedded in the general case, where the basic properties present themselves as simple, clear and 'practical'. It is precisely for this reason (and certainly not because of any dubious motive of wishing to start, come what may, by showing off, with no justification, the greatest generality and abstraction) that we did not begin the discussion with the case of events, and have still not stopped to consider it. Otherwise, we would have found ourselves, at this moment, having defined $\mathbf{P}(E)$ and not $\mathbf{P}(X)$, in more difficulty than if we had defined a unique concept, and with the unavoidable problem of producing $\mathbf{P}(X)$ as something not equally immediate, but as the combination of the $\mathbf{P}(E)$ and who knows what mathematical definition of integral.

3.6.2. *Some remarks concerning the two criteria*. Every operational definition, if one wants to take it too seriously as an actual method of measurement, carries with it the difficulty that the discussions of principle become mixed up with the doubts deriving from the practical imperfections inherent in any tool or procedure (these, however, often arise for reasons which may be important). Let us accept that this difficulty is unavoidable but that it is by no means a tragedy, since the definition deals with an idealized case, or limit-case, of conceptually possible experiments. Having said this, and having repeated that it is always infinitely better than any attempt at a mere verbal definition, emptily 'philosophical', there remains, nonetheless, the necessity of making oneself aware of the weak points in order to keep in mind the appropriate precautions.

We have already discussed, in Section 3.2, 'rigidity in the face of risk'; in other words, the temporary identification of utility and monetary value. At this juncture, a brief mention, with specific reference to the two criteria that we have put forward, will suffice. They both assume, implicitly, the hypothesis of rigidity. In the first place, they take the different bets, which are used as 'tests', to be summable; to be rigorous, the stipulation of any one of them should modify slightly the conditions for the stipulations of the others; secondly, by virtue of having a homogeneous character, in the sense that the procedure itself presupposes that $\mathbf{P}(aX) = a\mathbf{P}(X)$ (this adds no further restrictions, it is the same rigidity). This is useful if one attempts to limit the bets to be of moderate size; it is dangerous to allow it to be used indiscriminately. In the first criterion one has to think that in practice the opponent cannot impose excessively large bets (although the explicit inclusion of this kind of regulation in the definition would lead to a hybrid and tediously wordy exposition).

3.6.3. A defect of the first criterion is, in any case, the intervention of an 'opponent': this can make it difficult to avoid the risk, or at least the suspicion, of other factors intruding (such as the possibility of taking advantage of differences in information, competence, or shrewdness). By and large, such possibilities are in 'his' favour (he being the one who decides how much to bet, and in which direction; especially if he is the same person who has chosen the events for which the evaluation of probability is required). Sometimes, however, they can be in your favour (for instance, if, imagining the opponent to have a very distorted opinion, You enunciate an evaluation which induces him to stipulate a bet in a way that You judge favourable[20]).

3.6.4. Under the second criterion these negative features are not present (apart from that inherent in thinking of the various bets as summable). However, given that by choice of the coefficients k_i one can arrange the sizes of the penalties in whatever way one considers most appropriate, even this consequence of 'rigidity' becomes practically negligible. Observe, on the other hand, that the (arbitrary) choice of such coefficients – that is of the *metric* – has no influence at all on the implications of the criterion, since these are always based on merely *affine* notions: the notion of barycentre, and therefore its property of yielding a minimum for the moment, remains invariant under whatever metric one introduces for other purposes, and which occurs in the definition of the moment.

Another doubt arises: one might ask whether there is any good reason for considering the minimization of a penalty L, rather than the maximization of a prize $K - L$. Formally, there is no difference, but if one wants to fix K greater than any possible value of L (in order that the 'prize' always turns out to be positive) one is faced with an annoying limitation (which is impossible anyway if X is not bounded). There is, moreover, an historical reason: when introducing a similar theory in statistical applications, Wald found it natural to posit a Loss in the case of 'wrong decisions' (zero for correct decisions). Finally, one might exercise more care in attempting to prudently minimize a loss (which, in any case, involves uneasiness and disappointment), than in assuring oneself, in a reasonable manner, of the highest level of gain (in this context, the temptation to take a chance is often irresistible; naturally enough, since one cannot lose whatever happens).

3.6.5. One further remark, which is so deeply rooted in what we have said over and over again about the subjective meaning of probability that it is perhaps unnecessary. The two criteria are operational in the sense that they provide a means by which the opinions that an individual carries within himself, whatever they may be, turn out to be observable from the outside. There is no connection with questions like 'what is the *true* value of the probability?' – a question whose meaning finds no place within the present

20 On the eve of a certain football match, You attribute a probability of 40% to the victory of team A, but You think that your opponent, being a supporter of team A, evaluates it at 70%. You can then enunciate a probability of 65%, confident that he will hasten to pay 65 for that which to You is worth 40, but to him 70. But be careful! If he evaluates the probability at 50% instead, and decides to bet in the opposite way, he will pay You 35 for that which is worth 60 to You, and 50 to him – not 30 as You thought.

formulation, and whose meaning I was unable to discover in the attempts made by other theories to provide one – and not even with questions like 'how well founded, or how reasonable, are certain evaluations and their associated motivations?'.

This last question can, in a certain sense, be dealt with by reflecting on various problems that will present themselves from time to time as we proceed to study probability, and to examine various attitudes to both concrete applications and conceptual questions. However, it is mainly a question of arguments (of a rather psychological nature) concerning the choice of a single prevision, **P**, from among the infinite set of coherent previsions, $\mathscr{P}$ (which are equally acceptable from the mathematical point of view). The question does not concern mathematics, except in that it may give a still more enriched description of the various aspects of each choice, so that such a choice, always made absolutely freely, can be made by each individual after an accurate and straightforward examination of everything that in his personal judgment appears relevant for his decision.

3.7 Prevision in the Case of Linear and Nonlinear Dependence

3.7.1. Let us return to the two examples already introduced (Chapter 2, Section 10) under the guise of the logic of certainty. They lend themselves not only to illustrating in practice the application of the two criteria and the consequences of the properties we have established, but above all to developing necessary and instructive insights of a general character. First of all, one notes the essential connection between *linearity* and *prevision* (and the way in which this makes inapplicable to prevision certain arguments which would be valid for prediction). In this connection, it will become clearer how and why it is appropriate to extend the linear ambit $\mathscr{A}$ in relation to the questions to be examined (see the brief explanations given in Chapter 2, Section 2.8).

In the case of a ballot with n voters, we denoted by X, Y, Z, the number of votes cast in favour, against or abstaining, and considered in addition the difference and the ratio of votes for and against, which we denote by $U = X - Y$, $V = X/Y$.

If invited to make a prevision of the outcomes – on the basis of the first or second of the criteria put forward – you provide values $\bar{x}, \bar{y}, \bar{z}, \bar{u}, \bar{v}$. These are based on your knowledge, information, impressions or conjectures, about the inclinations or moods of the voters. If your values constituted a prediction, or if You intended to put them forward as *sure*, they would have to be chosen as *integers*, satisfying $\bar{x}+\bar{y}+\bar{z}=n$, $\bar{u}=\bar{x}-\bar{y}$, $\bar{v}=\bar{x}/\bar{y}$. In a prevision they might not even be integers: would the three relations hold? Is it valid to argue that they necessarily hold because they must hold for the true values? In fact, it might seem completely obvious that if the votes for and against *are* x and y (in reality, or in a prediction, or an estimation, or any prevision whatever) then their difference *is* $x - y$, and their ratio *is* x/y, whereas the number abstaining *is* $n - x - y$. For the two linear relations, given the linearity of **P**, this is certainly true, and, considering the ambit $\mathscr{A}$ of (x, y, u) – respectively, of (x, y, z) – is easily interpreted as follows: $\mathscr{Q}$ is the set of points with non-negative integer coordinates x and y with sum $\leqslant n$ in the plane $u = x - y$ – respectively, in the plane $z = n - x - y$ – and a barycentre of such points (with arbitrary weights) can only be some point in the given plane, in the triangle defined by $x \geqslant 0$, $y \geqslant 0$, $x + y \leqslant n$. If, however, we consider the $\mathscr{A}$ of (x, y, v), the points $\mathscr{Q}$ are those having the same x, y, but now on the surface (hyperbolic paraboloid) $v = x/y$, and the conclusion is no longer valid.[21]

In the case of a random prism with sides X, Y, Z, we denoted the diagonal by U, the area by V, and the volume by W. Since these are not linear functions of the sides, one would not expect that, having evaluated the previsions of the three sides as $\bar{x}$, $\bar{y}$, $\bar{z}$, those of the other elements, say $\bar{u}$, $\bar{v}$, $\bar{w}$, would satisfy the same relations as those holding between the true magnitudes. In other words, in the (six-dimensional) linear ambit $\mathscr{A}$ of (x, y, z, u, v, w), in which $\mathscr{Q}$ is the three-dimensional manifold with equations $u^2 = x^2 + y^2 z^2$, $v = 2(xy + yz + zx)$, $w = xyz$ (with x, y, z, u positive), one would not necessarily expect a barycentre of points of $\mathscr{Q}$ to lie in this manifold. It could be any point **P** whatsoever of $\mathscr{P}$, the convex hull of the given $\mathscr{Q}$: once we have evaluated $\bar{x}$, $\bar{y}$, $\bar{z}$, the previsions $\bar{u}$, $\bar{v}$, $\bar{w}$ can only turn out to be some point of the intersection of $\mathscr{P}$ with $x = \bar{x}$, $y = \bar{y}$, $z = \bar{z}$.

If one were interested in a complete solution to the problem, it would be necessary to determine $\mathscr{P}$, or this intersection with it. More generally, one should consider the same problem with certain restrictions on $\mathscr{Q}$: for instance, we might we aware of restrictions like $a' \leqslant X \leqslant a''$, $b' \leqslant Y \leqslant b''$, $c' \leqslant Z \leqslant c''$, or $d' \leqslant X + Y + Z \leqslant d''$, $X \leqslant Y \leqslant 2X$, $Y \leqslant Z \leqslant 2Y$, or that only integer values are admissible for X, Y, Z, or that there are only a finite number of values (x_i, y_i, z_i), and so on. For the purpose of illustration, a simpler version will do: let us suppose that Z is known, $Z = a$ say, and let us consider the restrictions that, given $\bar{x}$ and $\bar{y}$, result for $\bar{u}$, $\bar{v}$ and $\bar{w}$, *separately*, instead of jointly. In this way, everything is represented each time in a three-dimensional ambit $\mathscr{A}$, which is directly 'visible'.

In the ambit of (X, Y, U) the possible points $\mathscr{Q}$ lie on the circular hyperboloid $u^2 = a^2 + x^2 + y^2$; in fact, if there are no further restrictions, they are all the points on the 'quarter' $x \geqslant 0$, $y \geqslant 0$ of the sheet $u \geqslant 0$; otherwise, they are a subset of these. The barycentre of masses placed on this surface necessarily falls in the convex region that it encompasses (except in the trivial case where the mass is concentrated at a single point, a case where nothing is really random). In a coherent prevision, the diagonal must therefore *necessarily* be estimated longer than it would be if the lengths of the sides coincided exactly with their respective previsions. In the absence of other constraints, given $\bar{x}$ and $\bar{y}$, all the values lying between that minimum and $a + \bar{x} + \bar{y}$ are in fact admissible for $\bar{u}$: one approaches $a + \bar{x} + \bar{y}$ asymptotically by placing two small masses, $\bar{x}/k$ and $\bar{y}/k$, at the points $(k, 0)$ and $(0, k)$, with the rest at the origin, and then letting k become arbitrarily large.

In the ambit of (X, Y, W) the possible points $\mathscr{Q}$ lie instead on the hyperbolic paraboloid $w = axy$ (on the 'quarter' $x \geqslant 0$, $y \geqslant 0$). In the absence of other restrictions, the convex hull $\mathscr{P}$ is the entire positive orthant, since the barycentre can lie anywhere in this region. In other words, given $\bar{x}$ and $\bar{y}$, $\bar{w}$ can either coincide with $w = a\bar{x}\bar{y}$, or can be less (but bounded below by zero), or can be greater (with no constraint). The two limit-cases can be approached by simply placing all the mass at the origin, with the exception, in the first case, of two masses $\bar{x}/k$ and $\bar{y}/k$ at the points $(k, 0)$ and $(0, k)$, respectively, and, in the second case, a single mass $1/k$ at the point $(k\bar{x}, k\bar{y})$. The case $\bar{w} = a\bar{x}\bar{y}$ occurs,

21 In this example, provided we do not evaluate the probability that $Y = 0$ as *zero*, we actually have $\bar{v} = \mathbf{P}(V) = +\infty$. This is of more use in showing how one can encounter, in a natural way, cases where the hypothesis of boundedness does not hold, than in illustrating the proposition, which will be clearer after the following example.

for example, under a very important assumption – that of stochastic independence – which we are not yet in a position to discuss. The case of V reduces immediately (having set $Z = a$) to that of W; in fact, $2(XY + aX + aY) = 2W/a + 2a(X + Y)$.

The different behaviour in the two cases we looked at is due to that fact that the points of the first surface were all elliptic, always presenting the concavity in the same direction (delimiting in this way its convex hull), whereas for the second surface, whose points were all hyperbolic, the convex hull of each of its parts is necessarily formed of two parts, adhering to the two faces.

3.7.2. *Functional dependence and linear dependence.* In this context, the representation we have already introduced (Chapter 2, Section 2.8) by means of the dual spaces, $\mathscr{A}$ and $\mathscr{L}$, is appropriate, and provides some insight. $A(X)$ denotes the value which $X = f(X_1, X_2,..., X_n)$ would assume if the X_h (belonging to $\mathscr{L}$) assumed the values x_h (the coordinates of the point A of $\mathscr{A}$): in other words, $A(X) = f(x_1, x_2,..., x_n)$. This is only meaningful if A is one of the possible points $\mathscr{Q}$ of $\mathfrak{Q}$: that is the values x_h of the X_h are not incompatible. However, we now know that the other points in A – that is the points P of $\mathscr{P}$ the convex hull of $\mathscr{Q}$ – can be interpreted as previsions,[22] and one might ask whether $P(X)$ (understood as above, with $P = A$) is actually the prevision of X. It is clear that this only holds if X belongs to $\mathscr{L}$; in other words, if it is a linear function in the ambit $\mathscr{A}$, or, alternatively, if X is given not just by any function f of the X_h, but in fact by a linear function $X = \sum u_h X_h (h = 0,1,\ldots,n)$. The extension is only valid in this case, and that is why we always confine ourselves to using the notation $A(X)$. The above considerations give us another way of exhibiting the importance and the compass of the linearity. A point P (an admissible prevision) can be either a Q (that is a possible point) in the linear ambit $\mathscr{A}$, or a barycentre of possible points. The knowledge of the barycentre is sufficient, however, to determine only those things which remain invariant under any choice of the points Q and distribution of mass over them so long as one keeps the barycentre fixed.

In other words; one has always to recall that a **P**, defined on any set of random quantities X whatsoever (or, in particular, on any set of events $\mathscr{E}$), is uniquely extendible – and therefore defined – only on the linear space $\mathscr{L}$ of the (finite) linear combinations of $\mathscr{X}$; or, dually, in the corresponding linear ambit $\mathscr{A}$. If $\mathscr{L}$ (or $\mathscr{A}$) is enlarged, one can determine **P** more precisely by more or less arbitrary extensions. So long as we remain in a given ambit $\mathscr{A}$, each point **P** represents, in a manner of speaking, all the **P*** in some larger ambit which have **P** as their projection onto $\mathscr{A}$. This also holds in the infinite-dimensional case, but we postpone any explicit discussion until the Appendix. In order to clear up the simplest cases – one- or two-dimensions – it is sufficient to recall the examples already given in Section 3.7.1 and to examine these further aspects in that context.

3.7.3. *Conclusion.* We conclude, therefore, that whereas it is well known that coherent previsions *preserve* linear dependence, this *only* happens, in fact, *in this case.* In any other case it does not (unless by chance, or under suitable additional hypotheses) because the barycentre of masses lying in a given manifold need not itself belong to the

22 Moreover, it is possible to see that it can even be meaningful to consider an $A(X)$ where A does not belong to $\mathscr{Q}$, and not even to $\mathscr{P}$. For example, one might be interested in the difference between two $\mathscr{P}$, $A(X) = \mathbf{P}_1(X) - \mathbf{P}_2(X)$, and $A = \mathbf{P}_1 - \mathbf{P}_2$ certainly does not belong to $\mathscr{P}$ since we have $A(1) = 0$ instead of $A(1) = 1$.

manifold (except in the trivial case of linearity). In fact, it is sometimes impossible for it to do so (if $\mathscr{Q}$ is the boundary of a convex region).[23]

It is important to always bear in mind details of this kind and to reflect upon them. This is not only, and not mainly, because of their intrinsic importance – however notable this may be – but above all because one has to learn to free oneself from the ever present danger of confusing *prevision* and *prediction*. In a prediction any dependence should obviously be preserved (because it reduces to the choice of *a point* in $\mathscr{A}$, and not the barycentre of masses distributed over $\mathscr{Q}$). The type of argument which, in the examples given, turned out to be wrong if applied to prevision, would, on the other hand, be valid for a prediction. Despite knowing, and remembering, that the arguments do not hold for prevision, anyone (even You, even I) can inadvertently fall into error, applying them without sufficient thought in some particular problem, or in some small corner of the formulation of some particular problem.

There will be many and frequent occasions to warn against errors, misunderstandings, distortions, obscurities, contradictions and the other endless troubles that are so difficult to avoid when dealing with probability, and which are always essentially the result of ignoring the same warning: *prevision is not prediction*! It would not be a bad idea to imagine constantly in front of you an admonitory card – as is used by a certain well-known organization – bearing the message, 'Think!', but with an explanatory rider suited to the needs of probability theory and its applications:

$$\boxed{\textit{`Think : prevision is not prediction!'}}$$

There is an anecdote, concerning another such maxim, which may perhaps serve to make this recommendation more forceful. It reveals the fallacy of resorting to the self-deception of 'accepting for certain' the alternative on the basis of which one 'decides to act'; a vain attempt to replace a meaningful probability argument by an impossible translation of it into the inadequate logic of certainty. The anecdote is related by Grayson (on p. 52 of a book concerning which we shall have more to say: Chapter 5, 5.5.3) in the following way:

$$\boxed{\textit{Holes that are going to be dry shouldn't be drilled}}$$

'is printed on a sign hanging in one operator's office. This would truly be a "golden rule" if any oil or gas firm could live by it. Unfortunately, no one can – not even this particular operator who drilled 30 consecutive dry holes a few years ago.'

3.8 Probabilities of Events

3.8.1. The properties of probabilities of events are simply special cases of the properties of previsions of random quantities. It will be sufficient to establish them quickly, and to illustrate their meaning within the form of representation that we have introduced.

23 If we wished to be precise, we should exclude the points on the boundary where one does not have strict convexity: in other words, those which are barycentres of other points; or, alternatively, those through which there is no hyperplane which leaves *all* the other points of $\mathscr{Q}$ on the same side.

The theorem of 'total probability'. This is the name given to the theorem that translates, into the field of probability, the additive property of prevision.

The case of incompatibility. If two events A and B are incompatible then, as we have already noted, their logical and arithmetic sums coincide: $E = A \vee B = A + B$, so that, if **P** $\mathbf{P}(E) = \mathbf{P}(A) + \mathbf{P}(B)$. The same result holds for the (logical and arithmetic) sum of any finite number of incompatible events: $E = E_1 \vee E_2 \vee \ldots \vee E_n = E_1 + E_2 + \ldots + E_n$, and hence

$$\mathbf{P}(E) = \mathbf{P}(E_1) + \mathbf{P}(E_2) + \ldots + \mathbf{P}(E_n).$$

We can state this formally:

Theorem. *In the case of incompatible events, the probability of the event-sum must be equal to the sum of the probabilities.*

The case of (finite) partitions. In particular, for a partition in which, in addition, the sum $E = 1$, and hence $\mathbf{P}(E) = 1$, one has the following:

Theorem. *In a (finite) partition the probabilities must sum to* 1.

In particular, for two complementary events E and $\tilde{E}$ (a partition with $n = 2$), it turns out that $\mathbf{P}(E) + \mathbf{P}(\tilde{E}) = 1$; that is to say, $\mathbf{P}(\tilde{E}) = 1 - \mathbf{P}(E) = {\sim}\mathbf{P}(E)$; or, in yet another form, if $\mathbf{P}(E) = p$, then $\mathbf{P}(\tilde{E}) = \tilde{p}$.

In words:

Theorem. *The probabilities of two complementary events must themselves be complementary.*

Recalling the properties of the constituents, one can state immediately the following:

Corollary. *In order that the probabilities of all the events E which are linearly dependent on $E_1, \ldots, E_n$ should be determined, it is necessary and sufficient to attribute probabilities to all the constituents $C_1 \ldots C_s$. These probabilities must sum to* 1; *the* $\mathbf{P}(E)$ *depend linearly upon them.*

3.8.2. *Sufficiency of the conditions.* The preceding statements tell us how 'we must' – or 'You must' – evaluate probabilities; in other words, they impose necessary conditions for coherence. In fact – with the obvious restriction that the probabilities be non-negative – they are also sufficient, in the sense that an evaluation satisfying them is coherent, no matter how You choose it. We have already seen this in general in Section 3.4; it may be useful to repeat the argument in this particular case where it is very simple and clear.

Suppose that to the events $E_1 \ldots E_n$ of a finite partition You have attributed non-negative probabilities $p_1 \ldots p_n$, summing to 1, and that I (thinking in terms of 'The first criterion' of Section 3.3) try to force You into a bet which assures me of certain gain. I have to fix the amounts c_i for the bets on the individual E_i in such a way that the resulting bet

$$X = c_1(E_1 - p_1) + c_2(E_2 - p_2) + \ldots + c_n(E_n - p_n)$$

is certainly positive; in other words,

$$c_1E_1 + c_2E_2 + \ldots + c_nE_n > c_1p_1 + c_2p_2 + \ldots + c_np_n$$

no matter which of the E_i occurs. If E_i occurs, however, the left-hand side has the value c_i, and it is impossible for this to be always greater than the right-hand side, since the latter is itself a weighted average of the c_i.

3.8.3. *The case of compatibility; inequality.* For any arbitrary set of events, that is without making the assumption of incompatibility, we have

$$E = E_1 \vee E_2 \vee \ldots \vee E_n = 1 \wedge (E_1 + E_2 + \ldots + E_n) \leqslant E_1 + E_2 \ldots + E_n$$

and hence

$$\mathbf{P}(E) \leqslant \mathbf{P}(E_1) + \mathbf{P}(E_2) + \ldots + \mathbf{P}(E_n). \tag{3.6}$$

Stated formally:

Theorem. *The probability of the event-sum must be less than or equal to the sum of the probabilities.*

This is even more evident if one puts it in the form

$$\mathbf{P}(E) \leqslant \mathbf{P}(E_1 + E_2 + \ldots + E_n);$$

that is that the probability of the event-sum must be less than or equal to the prevision of the number of successes (one only has to consider that the latter takes into account multiplicities, whereas the former does not).

Expressions in terms of products. In the case of compatible events nothing can be said about $\mathbf{P}(E)$ other than the preceding inequality based just on the $\mathbf{P}(E_i)$. If we introduce other elements, and evaluate them, then, of course, things change. In terms of constituents, the only one we require is

$$C = \tilde{E}_1 \tilde{E}_2 \ldots \tilde{E}_n \quad \text{because} \quad E = \tilde{C}, \quad \mathbf{P}(E) = 1 - \mathbf{P}(C).$$

Making use of the products of the E_i (two at a time, three at a time, etc.), and the expansion

$$E = \sum_i E_i - \sum_{ij} E_i E_j + \sum_{ijh} E_i E_j E_h - \ldots \pm E_1 E_2 \ldots E_n, \tag{3.7}$$

we have at once the following:

Theorem. *For the probability of the event-sum we must always have*

$$\mathbf{P}(E) = \sum_i \mathbf{P}(E_i) - \sum_{ij} \mathbf{P}(E_i E_j) + \sum_{ijh} \mathbf{P}(E_i E_j E_h) - \ldots \pm \mathbf{P}(E_1 E_2 \ldots E_n). \tag{3.8}$$

Observe that the expression is linear in the probabilities of the products. Note also the special cases of two and three events:

$$\mathbf{P}(A \vee B) = \mathbf{P}(A) + \mathbf{P}(B) - \mathbf{P}(AB),$$
$$\mathbf{P}(A \vee B \vee C) = \mathbf{P}(A) + \mathbf{P}(B) + \mathbf{P}(C) - \mathbf{P}(AB) - \mathbf{P}(BC) - \mathbf{P}(AC) + \mathbf{P}(ABC).$$

3.8.4. *Extensions.* The same formula serves to express the probability that out of n events a given h occur, and no others; and hence the probability that exactly h occur (no matter which ones). The occurrence of $E_1E_2 \ldots E_h$, and no others, can be written as:

$$\begin{aligned} & E_1E_2\ldots E_h(1-E_{h+1})(1-E_{h+2})\ldots(1-E_n) \\ & = E_1E_2\ldots E_h - \sum_i E_1E_2\ldots E_hE_{h+i} + \sum_{ij} E_1E_2\ldots E_hE_{h+i}E_{h+j} - \ldots \pm E_1E_2\ldots E_n \end{aligned} \tag{3.9}$$

(where, as can be seen, the sum with k indices is the sum of the products of $h + k$ events; the given h together with k of the others). The event $Y = h$, the number of successes = h, is the sum of $\binom{n}{h}$ events of the above kind; in other words, the sum of all the corresponding expressions. In this sum, the products h at a time appear only once, those $h + 1$ at a time appear $h + 1$ times (once for each combination h at a time of their $h + 1$ factors), and so on; in general, the products $h + k$ at a time each appear $\binom{h+k}{h}$ times. For this reason, denoting the sum of the products r at a time by $\Sigma(r)$ for convenience, we have

$$\begin{aligned} (Y=h) &= \sum{}^{(h)} - \binom{h+1}{h}\sum{}^{(h+1)} + \binom{h+2}{h}\sum{}^{(h+2)} - \ldots \pm \binom{n}{h}\sum{}^{(n)} \\ &= \sum_{r=h}^{n}(-1)^{r-h}\binom{r}{h}\sum{}^{(r)}. \end{aligned} \tag{3.10}$$

If in place of the $\Sigma(r)$ we substitute the sum *of the probabilities* of the products, $\mathbf{P}(E_{i_1}E_{i_2}\ldots E_{i_r}) = p_{i_1i_2\ldots i_r}$, which we denote by S_r for short, the same formula gives the probability

$$\mathbf{P}(Y=h) = \sum_{r=h}^{n}(-1)^{r-h}\binom{r}{h}\sum p_{i_1i_2\ldots i_r} = \sum_{r=h}^{n}(-1)^{r-h}\binom{r}{h}S_r. \tag{3.11}$$

Note in particular:

$$\begin{aligned} &\mathbf{P}(Y=0) = 1 - S_1 + S_2 - S_3 + \ldots \mp S_{n-1} \pm S_n \\ &\mathbf{P}(Y=1) = S_1 - 2S_2 + 3S_3 - 4S_4 + \ldots \mp (n-1)S_{n-1} \pm nS_n \\ &\mathbf{P}(Y=2) = S_2 - 3S_3 + 6S_4 - 10S_5 + \ldots \mp \binom{n-1}{2}S_{n-1} \pm \binom{n}{2}S_n \\ &\qquad\qquad\qquad\ldots \\ &\mathbf{P}(Y=n-1) = S_{n-1} - nS_n \\ &\mathbf{P}(Y=n) = S_n \end{aligned}$$

(where $\pm$ stands for $(-1)^n$, and $\mp$ for $-(-1)^n$).

Example. A classical and instructive problem is that *of matching*, which lends itself to amusing formulations. If one has n letters and their respective envelopes, what is the probability that if *the letters are inserted into the envelopes at random* one has none, or one, or two, ..., or n '*matchings*'; that is letters in their own envelopes? The same problem arises if one pairs up at random right and left shoes from n pairs, or the husbands and wives of n couples, or the jackets and trousers of n suits, and so on. Alternatively, if one gives back at random to n people their passports, the keys of their hotel rooms, hats left in the cloakroom, and so on. More standard versions are given by the matchings in position among playing cards from two identical decks (for instance by placing them at random in two rows), or between the number of the drawing from an urn of numbered balls and the number of the ball drawn.

The probability of a matching at any given position is obviously $1/n$, of two matchings at two given positions is $1/[n(n-1)]$, and, in general, of r matchings at r given positions is

$$\frac{1}{\left[n(n-1)\ldots(n-r+1)\right]}=\frac{(n-r)!}{n!}$$

(in fact: only one out of the n objects, or only one out of the $n(n-1)$ pairs, …, or only one out of the $n!/(n-r)!$ arrangements r at a time, is favourable)

The S_r are therefore the sum of $\binom{n}{r}$ terms all equal to $(n-r)!/n!$, so that $S_r = 1/r!$ (independent of n), from which, denoting the number of matchings by Y, we obtain

$$\mathbf{P}(Y=0)=\left\{1-1+\frac{1}{2!}-\frac{1}{3!}+\ldots\pm\frac{1}{n!}\right\}=\mathrm{e}^{-1}-R_n\cong\mathrm{e}^{-1}$$ [24]

$$\mathbf{P}(Y=h)=\left\{1-1+\frac{1}{2!}-\frac{1}{3!}+\ldots\mp\frac{1}{(n-h)!}\right\}\Big/h!=\left(\mathrm{e}^{-1}-R_{n-h}\right)/h!\cong\mathrm{e}^{-1}/h!$$

(in particular: $\mathbf{P}(Y=n-1)=0$, $\mathbf{P}(Y=n)=1/n!$). Expressed numerically, $\mathrm{e}^{-1}=0{\cdot}367879$.

In the limit, as n increases, the distribution tends to that in which

$$\mathbf{P}(Y=h)=\mathrm{e}^{-1}/h!$$

(as we shall see later, Chapter 6, 6.11.2, this is the Poisson distribution with prevision $\mathbf{P}(Y)=1$).

Observe that for the matching problem one could establish immediately by a direct argument that $\mathbf{P}(Y)=1$ (i.e. that, in prevision, there is only one matching, whatever n is). We have only to note that it is given by $\mathbf{P}(Y)=n.(1/n)$, since the prevision (probability) of a matching at any one of the n places is $1/n$.

Observe also that the relation $\mathbf{P}(Y=n-1)=0$ is obvious: in fact, $n-1$ is not a possible value for Y because if we have matchings in $n-1$ positions the last one cannot fail to give a matching (it is as well to point out this fact since it is easily overlooked!).

3.8.5. *Entropy*. Given a partition into events with probabilities $p_1, p_2, \ldots, p_n$, we define the entropy to be the number

$$\sum_h p_h\left|\log_2 p_h\right|\quad\left(\log_2 p_h=(\log p_h)/(\log 2)\right),$$

24 R_n is the remainder of the series $\Sigma\pm 1/k!$ from the term $\pm 1/(n+1)!$ onwards: it is approximately equal to this first omitted term (which exceeds it in absolute value). With respect to e^{-1} it is practically negligible, except when n (respectively $n-h$) is very small (even for $n=10$ or $n-h=10$, the correction does not affect the decimal expression given for e^{-1}).

which represents the prevision of the number of YES–NO questions required to identify the true event.

This is immediate in the case of $n = 2^m$ equally probable events: m YES–NO questions are necessary and sufficient to know with certainty to which half, quarter, eighth,…, of the partition the true event belongs, and finally to know precisely which one it is. If we have nine events with probabilities $\frac{1}{12}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{32}, \frac{1}{32}, \frac{1}{32}, \frac{1}{64}, \frac{1}{64}$, one question suffices if the first one is true; if not (with probability $\frac{1}{2}$) another two questions are sufficient to decide which one of the next three is true, or whether the true event is one of the remaining five; finally (with probability $\frac{1}{8}$), another two questions are necessary, plus (with probability $\frac{1}{32}$) a further one to decide between the last two events. The entropy in this example is therefore given by

$$1+2\left(\tfrac{1}{2}\right)+2\left(\tfrac{1}{8}\right)+\left(\tfrac{1}{32}\right)=2\tfrac{9}{32}=2\cdot 28.$$

If it is not possible to proceed by successive halvings, some fraction is wasted (unless some device is available: for the time being, however, this brief introduction will suffice).

The unit of entropy is called a *bit* (contraction of '*binary digit*'): in the example above, the entropy was 2·28 bits; in the case of $1024 = 2^{10}$ equally probable cases it is 10 bits. For a given n, the entropy is maximized by an equipartition ($p_h = 1/n$): the reader might like to verify this as an exercise.

An item of information that leads to the exclusion of certain of the possible outcomes causes a decrease in entropy: this decrease is called the *amount of information*, and, like the entropy, is measured in bits (it is, in fact, the same thing with the opposite sign: some even call it *negative entropy*). We note that, for the time being, we are not in a position to provide a complete explanation of our assertion that an increase in information causes a decrease in entropy.

3.8.6. *Probability as measure or as mass*. In the set-theoretic interpretation of the events, it appears natural to think of probability – a non-negative, additive function taking the value 1 on the whole space – either as a *measure*, or as a *mass*.

The most widely used approach at the present time is the systematic identification of events as sets, and probability as measure (with all the advantages – as well as the risks! – that derive from a mechanical transposition of all the concepts, procedures and results of measure theory into the calculus of probability). To those reservations that we have already repeatedly expressed in connection with the systematic adoption of the set-theoretic interpretation of events, others must be added (in our opinion) concerning the further inflexibility introduced by the identification of probability with measure. This can, in fact, lead one to think that the representation in a space furnished with a measure binds events and random entities inseparably to a well-determined evaluation of probability. In the most elementary case, where we use Venn diagrams, the figures should be drawn in such a way that the area of each section be equal to its probability (taking the basic rectangle to have unit area). This, on the other hand, is in accordance with those points of view in which to each event (set) there corresponds an objectively (or, in any case, uniquely) determined probability.

If, instead, one wishes to distinguish between, on the one hand, the representation of the logical situation and, on the other hand, the introduction of whatever coherent evaluation of probability one wants to make, it turns out to be preferable to think of probability as *mass*. The mass can, in fact, be distributed at will, without altering the geometric support and the 'measure', which might in that context appear more natural.[25] In the Venn diagram, without changing the figure in any way, there is no difficulty in imagining the possible ways of distributing a unit mass among the various parts (it does not matter if we put large masses on small pieces) and in imagining those ways that various individuals, real or hypothetical, would have chosen as their own opinion, or those we think they might choose.

Another advantage is the following: if one gives to the space of the representation the structure of the linear ambit, $\mathscr{A}$, then the well-known implications of the mechanical meaning of *mass* make clear all those probabilistic properties that can be translated in terms of knowledge of the barycentre of a distribution (as we have already had occasion to see), or of moments of inertia, and so on.[26]

We shall see shortly some particularly significant applications of this concept in the linear ambit determined by n events (in the sense explained in Chapter 2, 2.8, where the 'possible points' are finite in number, since they correspond to the constituents). Meanwhile, before concluding these comments on the set-theoretic interpretation, it is perhaps instructive to point out the simple, but not obvious, meaning that the expression of the probability of the event-sum acquires under this interpretation. We will consider the case of three events, where

$$E = A \vee B \vee C = A + B + C - AB - AC - BC + ABC.$$

In the Venn diagram, Chapter 2, Figure 2.1b, the area of the union of the pieces A, B, C (or, alternatively, the mass contained in them) is calculated in the following way: firstly, we sum the areas of A, B and C; in this way, however, those of $AB\tilde{C}, A\tilde{B}C, \tilde{A}BC$ (doubly shaded) are counted twice, and that of ABC (triply shaded) is counted three times; subtracting those of AB, AC and BC, we re-establish the correct contribution for those originally counted twice; however, ABC is now counted three times less (since it belongs to AB and AC and BC) and therefore turns out to be ignored altogether; if we add it in, everything turns out as it should be.

3.9 Linear Dependence in General

3.9.1. The straightforward theorems concerning 'total probability', which we established at the beginning of the previous section, certainly require no further explanations. It is, however, convenient to introduce the use of the representation with the spaces $\mathscr{A}$ and $\mathscr{L}$ by means of the simple cases, before proceeding to others of a less trivial nature.

---

25 It is not that different distributions of 'mass' could not equally well be called different 'measures'. It is, however, a fact that when talking in terms of measure one tends to make of it something fixed, with a special status, whereas when talking in terms of mass there is the physical feeling of being able to move it in whatever way one likes.

26 The suggestion has even been put forward that one could always think just in terms of the mass (or measures, or area) rather than in terms of the original meaning of probability: in this way we avoid the questions and doubts of a conceptual nature to which such a notion of probability can give rise. In general, however, in addition to removing the doubts this would also remove the raison d'être of the problems themselves (unless these only involve formal aspects, capable of being isolated from the context which provides them with meaning and content).

We shall restrict ourselves, in general, to the three-dimensional case, which is the most obviously intuitive: the extension to n dimensions (which we shall occasionally mention) presents no difficulty for the reader who is familiar with such things, whereas for those who lack this familiarity it is better to be clear about the simpler case than to acquire confused and formal notions in a less accessible field.

Let E_1, E_2, E_3 be three events (which, for the moment, we take to be logically independent; we shall introduce various assumptions as we go on), and let (x, y, z) be the Cartesian reference system on which we superpose the linear ambit $\mathscr{A}$ and the linear space $\mathscr{L}$. The eight vertices of the unit cube

$$(0,0,0)(1,0,0)(0,1,0)(0,0,1)(0,1,1)(1,0,1)(1,1,0)(1,1,1),$$

thought of as points of $\mathscr{A}$, represent the constituents Q_i forming $\mathscr{Q}$;

$Q_0 =$	$Q_1 =$	$Q_2 =$	$Q_3 =$	$Q_1' =$	$Q_2' =$	$Q_3' =$	$Q_0' =$
$\tilde{E}_1\tilde{E}_2\tilde{E}_3$	$E_1\tilde{E}_2\tilde{E}_3$	$\tilde{E}_1E_2\tilde{E}_3$	$\tilde{E}_1\tilde{E}_2E_3$	$\tilde{E}_1E_2E_3$	$E_1\tilde{E}_2E_3$	$E_1E_2\tilde{E}_3$	$E_1E_2E_3$

(where negations correspond to the *zeros*, affirmations to the *ones*); thought of as points (or vectors) of $\mathscr{L}$, they represent the random quantities

$$0 \quad E_1 \quad E_2 \quad E_3 \quad E_2+E_3 \quad E_1+E_3 \quad E_1+E_2 \quad E_1+E_2+E_3$$

(where the presence of a summand corresponds to the *ones*).

The generic point (x, y, z), thought of as a point of $\mathscr{A}$, would mean that E_1 takes the value x, and similarly $E_2 = y$ and $E_3 = z$ (which is invalid, since the random quantities E_i cannot take on values other than 0, 1). This can be valid, however, as *prevision*, in the sense that $\mathbf{P}(E_1) = x$, $\mathbf{P}(E_2) = y$, $\mathbf{P}(E_3) = z$; in other words, (x, y, z) represents the prevision $\mathbf{P}$ which attributes to E_1, E_2, E_3 the probabilities $(p_1, p_2, p_3) = (x, y, z)$, and which is also expressible as the barycentre of the points Q_i with suitable weights (masses) q_i. Thought of as a point (or vector) of $\mathscr{L}$, (x, y, z) represents the random quantity $X = uE_1 + vE_2 + wE_3$ with coefficients $(u, v, w) = (x, y, z)$. Since $\mathbf{P}(X) = up_1 + vp_2 + wp_3 = ux + vy + wz$, $\mathbf{P}(X)$ can be interpreted as the inner product of the (dual) vectors $\mathbf{P}$ (or $P - 0$) of $\mathscr{A}$ and X (or $X - 0$) of L; or, alternatively, as $\mathbf{P}(X) = (P - 0) \times (X - 0)$ in the metric space on which $\mathscr{A}$ and $\mathscr{A}$ have been superposed.

Until we state precisely the assumptions made concerning the E_i, that is establish which among the eight products are actually possible constituents, all this remains rather general and introductory in character; simply a repetition of things we know already, with a few additional details.

3.9.2. *The case of partitions.* If the E_i constitute a partition, there are three constituents. $Q_1 = (1, 0, 0)$, $Q_2 = (0, 1, 0)$, $Q_3 = (0, 0, 1)$. We know that the p_i can be any three non-negative numbers summing to 1. In other words, the admissible $\mathbf{P} = (x, y, z)$ belong to the plane $x + y + z = 1$. More precisely, they belong to the triangle having as its vertices the three possible points Q_1, Q_2, Q_3, and are, in fact, uniquely expressible as barycentres of these points, $P = q_1Q_1 + q_2Q_2 + q_3Q_3$, with weights $q_1 = x$, $q_2 = y$, $q_3 = z$. This triangle constitutes the space $\mathscr{P}$ of admissible previsions, and is precisely the convex hull of the set $\mathscr{Q}$ of possible outcomes (which reduces in this case to the three given vertices). Representing the triangle by a figure in the plane, one sees that the probabilities x, y, z, turn out to be the barycentric coordinates of the point P with respect to the Q_i. Since the triangle is equilateral, one has the standard 'ternary diagram' (as is used, for example, to indicate the composition of ternary alloys) in which x, y, z also have a more immediate interpretation as the *distances of the point from the sides*, taking as unity the height of the triangle (to which the sum of the three distances is always equal). It is also clear that a point outside of the triangle (not in the plane, or in the plane but outside the triangle) can be brought nearer to all the three vertices – that is to all the points of $\mathscr{Q}$ – by transporting it into the triangle. This can be accomplished by projecting it onto the plane, and then, if the projection falls outside the triangle, by transporting it to the nearest point on the boundary. This is related to the 'second criterion', if we think of the penalty as being the square of the ordinary distance in this representation.

If we think in terms of $\mathscr{L}$, we could say, instead, that the point (1, 1, 1) represents the random quantity that is certainly equal to 1, given that $E_1 + E_2 + E_3 = 1$. The fact that for the coordinates of P we must have $x + y + z = 1$ is then interpreted on the basis of the scalar product: $\mathbf{P}(1) = x \cdot 1 + y \cdot 1 + z \cdot 1 = 1$.

3.9.3. *The case of incompatibility.* If the E_i are incompatible (but not exhaustive) there are four constituents: the previous three and $Q_0 = (0, 0, 0)$; that is Q_0, Q_1, Q_2, Q_3. The above considerations still hold, except that we now have the relation $x + y + z \leqslant 1$ (instead of = 1). We still have P expressible uniquely as a barycentre, $P = q_0Q_0 + q_1Q_1 + q_2Q_2 + q_3Q_3$, of the Q_i, with weights $q_0 = 1 - x - y - z$, $q_1 = x$, $q_2 = y$, $q_3 = z$, and the space $\mathscr{P}$(which was the triangle with vertices Q_1, Q_2, Q_3) is now the tetrahedron having in addition the vertex Q_0.

3.9.4. *The case of a product.* Let E_1 and E_2 be logically independent, and E_3 be their product: $E_3 = E_1E_2$. The constituents are then the following four: $Q_0 = (0, 0, 0)$, $Q_1 = (1, 0, 0)$, $Q_2 = (0, 1, 0)$ and $Q_0' = (1, 1, 1)$. The first three are in the plane $z = 0$, the last three are on $z = x + y - 1$; the other two groups of three are on $z = y$ and $z = x$, respectively. The space $\mathscr{P}$ is, therefore, the tetrahedron $z \geqslant 0$, $z \geqslant x + y - 1$, $z \leqslant x$, $z \leqslant y$, or, in other words, expressed compactly using $\wedge$ and $\vee$,

$$[\max(0,x+y-1)=] \qquad 0\vee(x+y-1)\leqslant z\leqslant x\wedge y \qquad [=\min(x,y)].$$

These are the restrictions under which one can arbitrarily choose the probabilities of two logically independent events and that of their product.

Here also, P is uniquely expressible as a barycentre

$$P = q_0Q_0 + q_1Q_1 + q_2Q_2 + q_0'Q_0'$$

of the Q with weights $q_0 = 1 - x - y + z$, $\quad q_1 = x - z$, $\quad q_2 = y - z$, $q_0' = z$.

3.9.5. *The case of the event-sum.* This proceeds as above, except that $E_3 = E_1 \vee E_2$ (instead of E_1E_2). Since the event-sum is $E_1 + E_2 - E_1E_2$, this case reduces straightaway to the preceding ones. The constituents are Q_0, Q_1', Q_2', Q_0'; the inequalities for the tetrahedron $\mathscr{P}$ having these vertices are

$$\Big[\max\big(x,y\big)=\Big] \qquad x\vee y\leqslant z\leqslant 1\wedge\big(x+y\big) \qquad \Big[=\min\big(1,x+y\big)\Big];$$

the weights which give $P = (x, y, z)$ as a barycentre in terms of the Q are

$$q_0 = 1-z, \qquad q_1' = z-y, \qquad q_2' = z-x, \qquad q_0' = x+y+z.$$

Remark. In the preceding cases each P was derived as a barycentre of the Q *with uniquely determined weights q;* it is important to note (and we shall return to this later) that this circumstance is exceptional. To be more precise, this happens when and only when the Q are *linearly independent* – in the examples above we had, in fact, either three noncollinear, or 4 noncoplanar – or when they are (as events) *expressible as a linear combination of the given events.* In fact, they were, in the first case, E_1, E_2, E_3; in the second, $1 - E_1 - E_2 - E_3, E_1, E_2, E_3$; in the third, $1 - E_1 - E_2 + E_3, E_1 - E_3, E_2 - E_3, E_3$; and in the fourth, $1 - E_3, E_3 - E_2, E_3 - E_1, E_1 + E_2 - E_3$. In other words, the Q (as events) belonged in these cases to $\mathscr{L}$. Observe that the expressions for the Q in terms of the E are the same as those for the weights q in terms of x, y, z. *In the following examples this will no longer happen.*

3.9.6. *The case of exhaustivity.* If we specify *only* that E_1, E_2, E_3 are exhaustive, then there are seven constituents; the eight minus $Q_0 = (0, 0, 0)$, which is excluded. This latter vertex of the cube being missing, the convex hull $\mathscr{P}$ is the cube itself minus the tetrahedron defined by this vertex and the three adjacent ones; that is the part of the cube $0 \leqslant x, y, z \leqslant 1$ which satisfies the inequality $x + y + z \geqslant 1$. Each of its points P can be expressed – *in an infinite number of ways* – as a barycentre of points Q (unless the point coincides with a vertex, or belongs to an edge, or a triangular face, in which case the number of representations is finite). In fact, all we have to do is to choose non-negative weights q, summing to 1, such that

$$q_1 + q_2' + q_3' + q_0' = x, \qquad q_2 + q_1' + q_3' + q_0' = y,$$
$$q_3 + q_1' + q_2' + q_0' = z$$

(4 equations and 7 unknowns).

3.9.7. *The case where the negations are also exhaustive.* If we exclude both the extreme constituents, that is in addition to $Q_0 = (0, 0, 0)$ we also exclude $Q_1' = (1, 1, 1)$, then six constituents remain. The cube has now had removed from it the two opposite tetrahedrons, and the remaining part $\mathscr{P}$ is that defined by the double inequality $1 \leqslant x + y + z \leqslant 2$.[27] Other considerations are as above.

A useful example is given by the comparisons between three random quantities, X, Y, Z; in other words, by considering the three events $E_1 = (X > Y), E_2 = (Y > Z), E_3 = (Z > X)$ (we assume excluded, or at least as practically negligible, the case of equality). By transitivity, the three events cannot turn out to be either all true or all false; there remain the other six constituents, corresponding to the $6 = 3!$ possible permutations. As an application, one might think, for example, of comparing the weights (or temperatures, etc.) of three objects.

Other cases. The following cases are similar (and are useful as exercises): $E_3 \subset E_1E_2$ (5 constituents); E_1 and E_2 incompatible, $E_3 \subset (E_1 = E_2)$ (6 constituents); and so on. Another example with four (independent!) constituents is given by $E_3 \equiv (E_1 = E_2)$.

3.9.8. *The case of logical independence.* All eight constituents exist; $\mathscr{P}$ is the whole cube. This is the most complete and 'normal' case; there is little to say apart from thinking about it in the light of remarks concerning more elaborate cases.

---

27 The case $x + y + z = 2$ (with constituents Q_1', Q_2', Q_3') is similar to that of the partition $(x + y + z = 1)$, and is obtained if $\tilde{E}_1, \tilde{E}_2, \tilde{E}_3$ form a partition.

The same case in any number of dimensions. If $E_1, E_2, \ldots, E_n$ are logically independent events, we will have 2^n constituents Q_i, the vertices of the unit hypercube; that is the points $(x_1, x_2, \ldots, x_n)$ in the linear ambit $\mathscr{A}$ with $x_i = 1$ or 0. The admissible previsions **P** are those of the cube $\mathscr{P}$, $0 \leqslant x_i \leqslant 1$, which is the convex hull of the set of the vertices $\mathscr{Q}$. The linear space L is formed by the random quantities $X = u_1E_1 + u_2E_2 + \ldots + u_nE_n$, which are linearly dependent (homogeneously, but it is easy to take into account separately an additive constant) on the events E_i. Conceptually, everything that has been stated for $n = 2$ and $n = 3$ also holds for arbitrary n (this saves us repeating everything in a more cumbersome notation and so making the exposition rather heavy going).

3.9.9. *General comments.* Each particular case differs from the final one by virtue of the exclusion of some of the constituents: instead of 2^n there are only $s < 2^n$. These determine a linear space of dimension $d(d \leqslant n, \log_2 s \leqslant d \leqslant s - 1)$; if $d < n$, the n events E_i are linearly dependent. In fact, if all the Q satisfy a linear relation $\sum_i x_i =$const. the same holds for the E_i. For instance, in the above examples $x + y + z = 1$, $x + y + z = 2$ gives $E_1 + E_2 + E_3 = 1$ (or 2), so that we need only consider two events, for example E_1 and E_2, setting $E_3 = 1 - E_1 - E_2$ or $E_3 = 2 - E_1 - E_2$, respectively (this also holds in the general case). If we consider the unnecessary E_i as eliminated (since they are linearly dependent on the others), we can always arrange that $d = n$; in any case, $\mathscr{P}$ is the convex hull (d-dimensional polyhedron) having as vertices the points Q which form $\mathscr{Q}$.

Given some **P** (in $\mathscr{A}$), in other words, *having evaluated the probabilities* $\mathbf{P}(E_i)$ *of the given events*, **P** turns out to be determined for all those random quantities X which are linearly dependent on the E_i, and for no others; that is for those belonging to $\mathscr{L}$. In particular, the probability of an event E is determined if and only if E is one of these X.[28] This statement takes into account all the obvious cases: for example, the probability of $A \vee B$ is not determined by $\mathbf{P}(A)$ and $\mathbf{P}(B)$ (unless we assume incompatibility), but is determined if we include $\mathbf{P}(AB)$, since we have the relation $A + B = AB + A \vee B$. It is useful to see an example of how nontrivial events can be found among the X of $\mathscr{L}$ (i.e. the X that only have two possible values; which we can always represent as 0 and 1). We shall see then that, if E is not linearly dependent on the E_i, one can only say that $p' \leqslant \mathbf{P}(E) \leqslant p''$, where $p' = \sup \mathbf{P}(X)$ for the X of $\mathscr{L}$ which are certainly $\leqslant E$, and $p'' = \inf \mathbf{P}(X)$ for the X of $\mathscr{L}$ which are certainly $\geqslant E$.

3.9.10 *A non-obvious example of linear dependence.* (This example is due to Lucio Crisma.) Suppose that A, B, C, D, F, G, H, K are the participants in a competition, and that six other individuals each choose among the participants their three 'favourites' (a prize being offered to all those who have included the winner among their 'favourites'). Suppose also that we know the choices to be: B, F, G for the first individual; A, D, K for the second; B, D, G for the third; B, G, H for the fourth; D, F, H for the fifth; C, G, K for the sixth. Finally, a seventh individual – suppose it is You – has chosen A, B, C.

28 This does not exclude the possibility that for certain evaluations (limit-cases in which an inequality reduces to an equality) $\mathbf{P}(E)$ can turn out to be determined for E which are not linearly dependent on the E_i we started with (and perhaps not even logically dependent). For instance, if neither of A and B is logically dependent on $A \vee B$, then knowing $\mathbf{P}(A \vee B)$ is not sufficient to determine $\mathbf{P}(A)$ and $\mathbf{P}(B)$; if, however, $\mathbf{P}(A \vee B) = 0$, it follows necessarily that $\mathbf{P}(A)$ and $\mathbf{P}(B)$ are also zero.

Is your guess, the event E, say, linearly dependent on the events $E_1, E_2, \ldots, E_6$, which denote the guesses of the others, or not? The question might be important, for example, in the following situation: there is an expert in whom You have great confidence, so far as judging the competition and the participants is concerned, and whose opinion concerning your probability of winning is of interest to You. However, you do not know this directly (since you do not know what probability of winning he attributes to each participant) but only indirectly (because you happen to know what probabilities he attributes to the guesses of the others turning out to be correct) : is this enough?

We have the system of equations:

$$\begin{aligned}
1 &= A+B+C+D+F+G+H+K\\
E_1 &= \quad B \qquad\quad +F+G\\
E_2 &= A \qquad +D \qquad\qquad +K\\
E_3 &= \quad B \quad +D \quad +G\\
E_4 &= \quad B \qquad\qquad +G+H\\
E_5 &= \qquad\qquad D+F \quad +H\\
E_6 &= \qquad C \qquad\quad +G \quad +K\\
E &= A+B+C.
\end{aligned}$$

One could verify that the six events $E_1, E_2, \ldots, E_6$ are linearly independent, but instead we note that we have the relation :

$$\begin{aligned}
&2E_1+3E_2-E_3+2E_4+4E_5+3E_6+3E\\
&=6(A+B+C+D+F+G+H+K)\\
&=6,
\end{aligned}$$

from which

$$E=\frac{1}{3}\left(6-2E_1-3E_2+E_3-2E_4-4E_5-3E_6\right).$$

Hence, if I know the $p_i=\mathbf{P}(E_i)$ of the guesses, I can conclude that in the expert's opinion (assumed coherent) $p=\mathbf{P}(E)$ must be

$$p=\frac{1}{3}\left(6-2p_1-3p_2+p_3-2p_4-4p_5-3p_6\right).$$

3.10 The Fundamental Theorem of Probability

3.10.1. We turn now to proving and illustrating the general conclusion that we stated before, and which, in a more complete and precise form, constitutes the following:

Theorem. *Given the probabilities* $\mathbf{P}(E_i)$ $(i = 1, 2,\ldots, n)$ *of a finite number of events, the probability,* $\mathbf{P}(E)$, *of a further event E, either*

a) *turns out to be determined (whatever* $\mathbf{P}$ *is) if E is linearly dependent on the* E_i *(as we already know); or*
b) *can be assigned, coherently, any value in a closed interval* $p' \leqslant \mathbf{P}(E) \leqslant p''$ *(which can often give an illusory restriction, if* $p' = 0$ *and* $p'' = 1$*, or in limit-cases for particular* $\mathbf{P}$*, give a well-determined result* $p = p' = p''$*).*

More precisely, p' *is the upper bound,* sup $\mathbf{P}(X)$, *of the evaluations from below of the* $\mathbf{P}(X)$ *given by the random quantities X of* $\mathscr{L}$ *(i.e. linearly dependent on the* E_i*) for which we certainly have* $X \leqslant E$*. If E is not logically dependent on the* E_i*, observe that* $X \leqslant E$ *can be more usefully replaced by* $X \leq E'$ *where* E' *is the largest event logically dependent on the* E_i *contained in E (see Chapter 2, 2.7.3). The same can be said for* p'' *(replacing* sup *by* inf, *maximum by minimum,* E' *by* E''*, and changing the direction of the inequalities, etc.).*

Proof. If $Q_1 \ldots Q_s$ denote the constituents, relative to $E_1 \ldots E_n$, and E is logically (but not linearly) dependent on the E_i, then the linear ambit $\mathscr{A}'$ obtained by the adjunction of E (i.e. by adding a new coordinate x to the preceding $x_1 \ldots x_n$) has the same constituents Q_h, but now placed at the vertices of a cube in $n + 1$ dimensions instead of n. Each $Q = (x_1, x_2,\ldots, x_n)$ is either left as it was (with $x = 0$), or moved onto the parallel S_n $(x = 1)$, becoming either $(x_1, x_2,\ldots, x_n, 0)$ or $(x_1, x_2,\ldots, x_n, 1)$, according to whether Q is contained in $\tilde{E}$ or in E. The convex hull $\mathscr{P}'$ in S_{n+1} (in $\mathscr{A}'$) has as its projection onto the preceding S_n ($\mathscr{A}$) the preceding $\mathscr{P}$. For each admissible $\mathbf{P}$ in the latter (with coordinates $p_i = \mathbf{P}(E_i)$), the admissible extensions in $\mathscr{A}'$ are the points $\mathbf{P}'$ that project onto $\mathbf{P}$ and belong to $\mathscr{P}'$; that is. belong to the segment $p' \leqslant x \leqslant p''$ which is the intersection of the ray $(p_1, p_2, \ldots, p_n, x)$ with $\mathscr{P}'$. The extreme points $(x = p'$, $x = p')$ are on the boundary of $\mathscr{P}'$, that is on one of the hyperplanes (in n dimensions) that constitute its faces (they could be on more than one – vertices, edges, etc. – but this does not affect the issue). Suppose the hyperplane is given by $\sum u_i x_i + ux = c$; in other words, suppose that the relation $\sum u_i E_i + uE = c$ holds on it, that is that $E = (c - \sum u_i E_i)/u$: then the X in $\mathscr{L}$ defined by the right-hand side has the given property, and yields $p' = \mathbf{P}(X)$. Similarly for p''.

3.10.2. *Applications.* Let us generalize some of the examples considered previously in S_3. Those concerning the number of successes,

$$Y = E_1 + E_2 + E_3,$$

now become the consideration of $Y = E_1 + E_2 + \ldots + E_n$, and we can look at various sub-cases. Suppose that either Y is known, $Y = y$ $(0 \leqslant y \leqslant n)$ (as in the previous cases where $Y = 1$ and $Y = 2$), or certainly lies between two given extreme values y' and y'' $(0 \leqslant y' \leqslant y'' \leqslant n)$ (as in the previous cases, where $1 \leqslant Y \leqslant 2$). The interpretation of this last example, as given in Section 3.9.7, will now be extended (in different ways) to comparisons between n objects: finally, the case of the event-sum will require all the products.

3.10.3. *Knowledge about frequency.* This first example is noteworthy in that it constitutes the first and most elementary link in the long chain of conclusions which, as we proceed, will clarify and enrich our insight into the relationship that holds between *probability* and *frequency*. This is important both for what the conclusions do say and, perhaps even more so (in some situations at least), in order to get used to not interpreting them as saying something which they do not say.

The simplest case is that in which the number of successes, $Y = E_1 + E_2 + \ldots + E_n$, is known (for certain);that is the *frequency* Y/n is known (for certain). Let $Y = y$, so that $Y/n = y/n$. The following are possible examples: in an election, out of n candidates we know that y are to be elected; in an examination, y candidates out of n passed (but we are still ignorant of which ones); in a drawing of the lottery, out of $n = 90$ numbers $y = 5$ will be drawn; at $n = 90$ successive drawings of all the balls in Bingo, all the $y = 15$ numbers on your card will come out.

As an extension, we have the case in which we know the limits between which Y must lie; $y' \leqslant Y \leqslant y''$ (and hence that the frequency must be between y'/n and y''/n). In the preceding examples: it may be that the electoral system allows the number elected to vary between y' and y''; that on the basis of partial information about the examinations one knows that at least y' have passed and at least $n - y''$ have not; if we consider 10 drawings of the lottery instead of one (for instance, all the 10 'wheels' on the same day), then of the $n = 90$ numbers the total of different numbers drawn can vary between $y' = 5$ (all the sets of five identical) and $y'' = 50$ (no number repeated).

It is obvious that, as in the case $n = 3$, the sum of the $\mathbf{P}(E_i)$, that is $\mathbf{P}(Y)$, must give in the first case y, and in the second a value $y' \leqslant \mathbf{P}(Y) \leqslant y''$. Put more forcefully; dividing by n, *the probabilities* $\mathbf{P}(E_i)$ *must be such that their arithmetic mean coincides with the known frequency* y/n, *or falls between the extreme values*, y'/n *and* y''/n, *that the frequency can assume* (end-points included). This is all that can be said on the basis of the given information. In general, one might say more: for example, that each number in the lottery has probability $\frac{5}{90}$ of coming up in a given drawing, and not different probabilities with mean $\frac{5}{90}$. This could only be done, however, on the basis of additional knowledge or considerations which must be kept separate.

3.10.4. *The linear ambit of events logically dependent on n given events.* For the purpose in hand, it is obviously sufficient to consider the linear ambit, let us call it $\mathscr{A}^*$, generated by the s constituents Q_h (these form a partition, and so the dimension is actually $s - 1$, given the identity $Q_1 + Q_2 + \ldots + Q_s = 1$). We could also generate it by means of the E_i and their products (two at a time, three at a time, etc.). We saw, in Section 3.8.3, that in this way one can express the event-sum linearly, and we shall now see that it is possible to express all the constituents linearly, and hence all the events which are logically dependent on the E_i. We will suppose that the E_i are logically independent, so that $s = 2^n$; in the other case, the treatment is equally valid, except that the constituents and the products which turn out to be impossible have to be omitted.

Let us illustrate the situation by referring to the case of three logically independent events and their products; for convenience we denote the three events by A, B, C (instead of E_1, E_2, E_3) and their products by $F = AB$, $G = AC$, $H = BC$ and $E = ABC$. We have seven events that are linearly independent because there exists only one linear relation between the $2^3 = 8$ constituents (the sum $=1$). Some inequalities (implications) hold among them, however; for instance, $A \geqslant AB \geqslant ABC$ so that $A \geqslant F \geqslant E$ (as is obvious if one considers that of the $2^7 = 128$ vertices of the cube in seven dimensions only the eight corresponding to the constituents relative to A, B, C, are possible).

We list the constituents, giving their coordinates in the ambit $\mathscr{A}^*$, and the linear expressions in the dual space $\mathscr{L}^*$:

$$
\begin{aligned}
&ABCFGHE=(1,1,1,1,1,1,1)=E,\\
&AB\tilde{C}F\tilde{G}\tilde{H}\tilde{E}=(1,1,0,1,0,0,0)=F-E,\\
&A\tilde{B}C\tilde{F}G\tilde{H}\tilde{E}=(1,0,1,0,1,0,0)=G-E,\\
&\tilde{A}BC\tilde{F}\tilde{G}H\tilde{E}=(0,\ 1,\ 1,\ 0,\ \ 0,\ 1,\ 0)=H-E,
\end{aligned}
$$

$$
\begin{aligned}
&A\tilde{B}\tilde{C}\tilde{F}\tilde{G}\tilde{H}\tilde{E}=(1,\ 0,\ 0,\ 0,\ 0,0,\ 0)=A-F-G+E,\\
&\tilde{A}B\tilde{C}\tilde{F}\tilde{G}\tilde{H}\tilde{E}=(0,\ 1,\ 0,\ 0,0,0,\ 0)=B-F-H+E,\\
&\tilde{A}\tilde{B}C\tilde{F}\tilde{G}\tilde{H}\tilde{E}=(0,\ 0,\ 1,\ 0,\ 0,\ 0,\ 0)=C-G-H+E,\\
&\tilde{A}\tilde{B}\tilde{C}\tilde{F}\tilde{G}\tilde{H}\tilde{E}=(0,\ 0,\ 0,\ 0,\ 0,\ 0,0)=1-A-B-C+F+G+H-E.
\end{aligned}
$$

These expressions, and the analogous ones for each of the events logically dependent on A, B, C, are obtained as shown in the following example:

$$\tilde{A}B\tilde{C}=(1-A)B(1-C)=B-AB-BC+ABC=B-F-H+E.$$

The necessary and sufficient condition for coherence is that the probabilities of the constituents are non-negative (they automatically turn out to sum to 1), and therefore the following inequalities (where, for simplicity, we denote the probability of an event by the corresponding lower case letter) are necessary and sufficient:

$$e \geqslant 0, \qquad f, g, h \geqslant e, \qquad a \geqslant f+g-e, \qquad b \geqslant f+h-e,$$
$$c \geqslant g+h-e, \qquad (a+b+c)-(f+g+h)+e \leqslant 1.$$

3.10.5. *A canonical expression for random quantities*. By analogy, we indicate here how, in the same manner, each random quantity

$$X=c_0+c_1E_1+c_2E_2+\ldots+c_nE_n,$$

linearly expressible in terms of the events E_i, can be put in a meaningful canonical form by reducing it to a linear combination

$$X=x_1C_1+x_2C_2+\ldots+x_sC_s$$

of the constituents C_h (the x_h are the possible values of X, assumed in correspondence to the occurrence of the C_h). As an example: if we denote two logically independent events by A and B, and the constituents by $Q_1 = AB$, $Q_2 = A\tilde{B}$, $Q_3 = \tilde{A}B$, $Q_4 = \tilde{A}\tilde{B}$, where $1 = Q_1 + Q_2 + Q_3 + Q_4$, $A = Q_1 + Q_2$, $B = Q_1 + Q_3$, we have, for instance, for $X = 3 - 4A + B$:

$$
\begin{aligned}
X&=3(Q_1+Q_2+Q_3+Q_4)-4(Q_1+Q_2)+(Q_1+Q_3)\\
&=0\cdot Q_1+(-1)\cdot Q_2+4\cdot Q_3+3\cdot Q_4.
\end{aligned}
$$

X assumes the possible values −1, 0, 3, 4, corresponding to Q_2, Q_1, Q_4, Q_3.

3.10.6. *Comment*. The above considerations are intended to familiarize the reader (in the case of events) with the crucially important idea of the relations of linearity and inequality, and to stress *a fact* and *a criterion* that will be of use in what follows, and more generally.

The *fact* is the *possibility* of expressing all that can legitimately be said by arguing solely in terms of the events (and random quantities) whose prevision is known. That is to say, without leaving the linear ambit determined by the latter, without imagining already present a probability distribution over larger ambits, those in which the extension is possible, albeit in an infinite number of ways.

The *criterion* lies in the *commitment* to systematically exploiting this fact; the commitment considered as the expression of a fundamental methodological need in the theory of probability (at least in the conception which we here maintain). All this is not usually emphasized.

These considerations should go some way to excusing the length of the exposition, which is certainly excessive in comparison with what would be desirable if this topic were well enough known in general to permit us to restrict ourselves to a few brief remarks.

3.10.7. *The case of an infinite number of events* (*or random quantities*). The fundamental theorem of probability (and prevision), given in Section 3.10.1, permits us – even in countably infinite or nondenumerable cases where, of course, the number of choices is infinite – to proceed to attribute to all the events and random quantities that we wish, one after the other, probabilities and previsions coherent with the preceding ones. The arguments presented do not become invalid when we pass to the infinite case, because the conditions of coherence always refer just to finite subsets: see Appendix, Section 15.

This demonstrates the theorem of the *unconditional existence and extendibility of coherent previsions* of events and random quantities in any (open)[29] field. In other words:

> *If within the field in which they are made, the previsions do not already give rise to incoherence, no incoherence arises to prevent the existence of coherent previsions in any field whatever, coinciding with the preceding ones whenever these apply.*

3.11 Zero Probabilities: Critical Questions

3.11.1. In both the criteria put forward in order to define probability there was a point whose clarification we held over to the sequel. It was the same point in both cases; the wherefore of the precaution taken in excluding the possibilities of gains being *all uniformly negative*, but not that of gains being all negative (without the 'uniformity' condition). Another matter, connected with this, is the removal of the reservations regarding the prevision of unbounded random quantities.

We are dealing with critical questions and, if we only wished to consider those aspects relating to applications, they could be omitted, or confined to the Appendix. This, however, is not possible. In Chapter 6, we have to study distributions, and to throw light on the conceptual differences and their wherefores, introduced in accordance with the

29 'Open' is meant in the sense of not being preconstituted, not constrained, not a 'Procrustean bed', not a Borel field, not consisting of events, etc., that have a given meaning or structure, but a field in which we can, at any moment, insert whatever might come to mind.

present viewpoint, it is better to focus right from the beginning on those aspects which will play a fundamental rôle.

The fact is that a logical construction is such in so far as it is a whole in which 'tout se tient' ['everything fits together']; otherwise, it is nothing of the sort. Questions that are seemingly completely otiose and insignificant can have, and do have, interconnections with all the rest and are essential for an understanding of them. To ignore them, or merely to mention them in passing, is dangerous, especially when they impinge on delicate and controversial matters: too many ideas then remain rather vague and give rise to an accumulation of doubts.

For this reason, having reached the end of Chapter 3, we shall now consider the questions of a critical nature that have arisen; we shall do the same at the end of Chapter 4, coming back to these same questions under a new guise; finally, at the end of Chapter 5, we shall arrive at the same kind of considerations, although with respect to topics that are less technical and more general. We shall attempt to confine ourselves to the minimum necessary discussion, expressed as simply as possible. The few additional clarifications or examples will be recognizable 'at a glance' by virtue of the small print.

3.11.2. It would not be accurate to say that all the problems reduce to the presence of zero probabilities but, in order to have a guideline to follow, it is convenient to think in these terms (just as it is not only suggestive but also appropriate to mention them in the section heading).

It seems impossible that there is anything at all to be said about zero probabilities. Instead, we have the following basic questions:

i) Can a possible event have zero probability? If so:
ii) Is it possible to compare the zero probabilities of possible events (to say if they are equal, or what their relation is, etc.)?
iii) Can a union of events with zero probabilities have a positive probability (in particular, can it be the certain event)?
iv) Are there any connections with problems concerning random quantities, and in particular with the problem of prevision for unbounded random quantities?

Question (II) crops up again within the topics of Chapter 4 and will be discussed there; we had to mention it, not only to put it in its natural position as a 'question' but also to give prior warning that any incidental comments that we make here for convenience will be clarified at the appropriate place: we will draw attention to this by writing '(II!)'.

Questions (I) and (III) can be bracketed and discussed together straightaway; afterwards we shall pass on to (IV). However, there was a reason for putting the two questions (I) and (III) separately. Question (III), which evidently requires to be put in the context of infinite partitions, might lead one to think and state that one can only have possible events with zero probability *if they belong to infinite partitions* (!). This is monstrous. If E has probability p (in particular = 0) it *is* an event with probability p (in particular with zero probability) both when considered in itself, or in the dichotomy E and $\tilde{E}$, or in any other partition into few, many or an infinite numbers of events, obtained by partitioning in any way whatsoever. Unfortunately, this propensity to see each event embedded in some scheme, together with others usually studied with it, gives rise to serious confusions both in theoretical matters (as is the case here) and practically (as in the examples in Chapter 5, 5.8.7).

This having been said as an appropriate warning, we can pose question (III) once again by asking *whether in an infinite partition one can attribute zero probability to all the events*. In this form, the question becomes essentially equivalent to that concerning the different types of additivity: *finite*, only for a finite sum; *countable*, for the denumerable case; *perfect*, if the additivity always holds.

There are precisely three answers, corresponding to these three types (with a variation, which is related to (I)):

A = Affirmative, N = Negative (N' and N''), C = Conditional

(and in what follows, we shall denote them and the corresponding points of view with the initials A, N and C, or, if necessary, A, N', N'' and C).

A: Yes. Probability is finitely additive. The union of an infinite number of incompatible events of zero probability can always have positive probability, and can even be the certain event.

N: No. Probability is perfectly additive. In any partition there is a finite, or countable, number of events with positive probabilities, summing to one: the others have zero probability both individually and together.

C: It depends. The answer is NO if we are dealing with a countable partition, because probability is countably additive; the sum of a countable number of zeroes is zero. The answer is YES if we are dealing with an uncountable infinity,[30] *because probability is not perfectly additive: the sum of an uncountable infinity of zeroes can be positive.*

In the case of the answer N, there are, however, two subcases to be distinguished with reference to question (I) (for which, in cases A and C, the answer can only be YES).

N′: Probability zero implies impossibility. What has been said above is a consequence of this identification.

N″: Probability zero does not imply impossibility. However, the behaviour is the same: even if we take the union of them all, the events of probability zero form an event with zero probability.

3.11.3. Let me say at once that the thesis we support here is that of A, *finite additivity*; explicitly, *the probability of a union of incompatible events is greater than or equal to the supremum of the sums of a finite number of them.* Apart from the present author, it would seem that only B.O. Koopman (1940) has systematically adopted and developed this thesis. Others, like Good (1965), admit only finite additivity as an axiom, but do nothing to follow up this observation. Others again, like Dubins and Savage (1965), make use of finite additivity for special purposes and topics.

The thesis N is supported, as far as I know, only by certain logicians, such as Carnap, Shimony and Kemeny (as a consequence of a definition of 'strict coherence').[31]

30 I do not know whether this corresponds exactly to the conception of the supporters of this thesis (often one only talks about the case of the continuum).

31 In addition to these serious authors, there is no point in mentioning the large number who refer to zero probability as impossibility, either to simplify matters in elementary treatments, or because of confusion, or because of metaphysical prejudices.

The thesis C is the one most commonly accepted at present; it had, if not its origin, its systematization in Kolmogorov's axioms (1933). Its success owes much to the mathematical convenience of making the calculus of probability merely a translation of modern measure theory (we shall say a lot more about this in Chapter 6). No-one has given a real justification of countable additivity (other than just taking it as a 'natural extension' of finite additivity); indeed, many authors do also take into account cases in which it does not hold, but they consider them separately, not as absurd, but nonetheless 'pathological', outside the 'normal' theory.

3.11.4. Let us review, briefly, the main objections to the various theses (we number them: $A1$, $A2$,…; $N1$, $N2$,…; $C1$, $C2$,…). Our point of view is, of course, represented by the objections to N and C, and by the answers ($A1a$, $A1b$,…; $A2a$, $A2b$,…) to the objections raised against A. We will also interpolate some examples ($E1$, $E2$,…).

$A1$ This is an objection from the standpoint of N (or rather N'): *it is not sufficient to exclude as inadmissible those bets with gain X certainly negative* ($\longmapsto X < 0$: *weak coherence*); *it is necessary to exclude them if the gain is certainly nonpositive* ($\longmapsto X \leqslant 0$: *strict coherence*). *This means that 'zero probability' is equivalent to 'impossibility'.*

The most decisive reply will be objection $N2$, but it is better not to evade a reply that clarifies the points (perhaps persuasive) put forward in $A1$; this reply will constitute a preliminary refutation of N ($N1$).

$A1a$ It should be unnecessary to point out that the inadmissibility of a bet is always relative to the set of choices offered by a given scheme. It is obvious that if among the possible choices there was the choice 'do not make a bet at all', nobody would choose an alternative that could only lead to losses (this, however, means nothing).

$A1b$ In the simplest scheme, let $X = -E$ (loss = 1 if E occurs; e.g. the risk we are facing), and consider the appropriateness of insuring oneself by paying a premium p. Let us suppose that one is willing to pay $\left(\frac{1}{2}\right)^n$ (and no more) if $p = \left(\frac{1}{2}\right)^n$; for example E = all heads in n tosses. If E = all heads in an infinite number of tosses, I will not be willing to pay more than zero (every $\varepsilon > 0$ is $\left(\frac{1}{2}\right)^n$ for sufficiently large n, and would be too much even if the risk were infinitely greater). The lesser evil, therefore, is not to insure oneself; in other words, to act in this respect (*but not in others*) as if E were impossible.

$A1c$ There is more, however. The condition of coherence is and must be (as we established in Sections 3.3.5 and 3.3.6) *even weaker*[32] than the one criticized in $A1$, allowing in addition bets in which one can *only lose*! Let us suppose that an individual is subjected to a certain loss of a sum $1/N$ (where N is an 'integer chosen at random', with equal – and therefore zero – probabilities for each value, and hence for each finite segment $N \leqslant n$ (II!)). There is no advantage in paying a sum ε (however small) to avoid this certain loss, because it would always be practically certain that the loss avoided would be very much smaller.

$N1 = A1d$ Summarizing and concluding, we have the following. The variants (from the weakest to the strongest) consist in excluding X if

$$\sup X < 0, \quad \sup X \leqslant 0, \text{ with } X = 0 \text{ impossible}, \sup X \leqslant 0,$$

Objection $A1$ criticizes the middle statement, and supports the last one. In A1c we explained why, on the contrary, we think it necessary to support the first one.

32 If we wished to give this condition a name, we might call it *sufficient coherence* (in contrast to *weak* and *strict* coherence).

*N*2 *The variant N′ is logically absurd unless one excludes the possibility of considering a partition with an uncountable infinity of possible cases* (e.g. the continuum). In the denumerable case objections arise which also apply to *C* (*C*3 = *N*4, and so on).

*N*3 The variant *N″* does away with *N*2: nevertheless, *the meaning of zero probability is still exceptionally restrictive* (much more so than in *C*, and even there it is too restrictive; see *C*4).

In fact, one should be able to define E^* = the union of all events with zero probabilities = the maximal event with zero probability (let us call it 'the catastrophe'). Under the noncatastrophic hypothesis (with probability = 1) one goes back to *N′*; only in the opposite cases are events with zero probability no longer impossible (and, consequently, (II!) can have any probability whatsoever).

3.11.5. *C*1 *C appears to be less logically plausible than A and N* – we suspect 'Adhockery for mathematical convenience' – because the distinction between finite and infinite has without doubt a logical and philosophical relevance, whereas it might seem strange to draw the crucial distinction between finite and nondenumerable on the one hand, and countable on the other hand.

*C*2 A difficulty that derives from this is the following: *given a partition* (*e.g. whose cardinality is that of the continuum*) *into events of zero probability*, what happens if as a consequence of additional information one believes that *only a countable infinity remain possible*? In particular, if one assumes them (II!) equally probable? Or under the most general hypothesis?

*E*1 Initially, *X* has a uniform distribution over the real numbers between 0 and 1 (all points equally probable (II!)). Additional information reveals that *X* is rational.

*E*2 It seems obvious (but recall (II!)) that in this case – that is *E*1 after the given 'additional information' – *the values which remain possible, that is the rational values of* [0, 1], *are* (*still*) *equally probable* (they define a 'random choice' from the original set).

If one thought of actually interpreting the problem geometrically, one might perhaps doubt the judgment of all the rationales as equally probable, considering as 'rather special' the end-points, mid-point, fractions with small denominator, decimal fractions with only a few figures and so on.

This effect is lessened if one thinks of taking the 'distance between two points chosen at random' (the first minus the second; if negative add 1, take the result mod 1).

It disappears altogether if one thinks in terms of a circle obtained by rolling up the segment without indicating which is the 'zero' point.

*C*3 = *N*4 Objection *C*2 can also be raised in the countable case (and then it also concerns *N*). *Suppose that we have a countable infinity of possible cases, one with* $p = 1$ (*and the others therefore with* $p = 0$); *assume we know* that *the first one has not occurred*.

*E*3 Let *N* be the number of passages through the origin in a random walk for which $\mathbf{P}(N > n) = 1$ for all *n* (an example is Heads and Tails); information: $N \neq \infty$.[33]

---

33 This information could only be given by somebody who had explored the world as it appears after the end of time.... Objections to 'lack of realism' would, however, be out of place here as it is merely a question of logical compatibility. Where they are appropriate (and usually insufficiently dealt with), the exigencies of realism will be examined here (especially in the Appendix), perhaps at greater length than hitherto, and perhaps more than is reasonable. One cannot refute the exact nature of a conclusion based on the examination of a 'pathological' curve (e.g. that of Helge van Koch) by the pretext that there exist neither pencils, nor sheets of paper, nor hands, by means of which it could be drawn.

*E*4 In general, in such cases it is plausible to say that the

$$p_h = \mathbf{P}(N = h \mid N \neq \infty)$$

are all zero, and (II!) each is infinitely greater than the preceding one. We limit ourselves to a mere statement of this in order to be able to refer to this example without examining it deeply.

*C*4 *The meaning of* $p = 0$ *is too restrictive even in C* (although much less so than in *N*; see *N*3). Expressed in a vague form, but one which corresponds exactly to the state of things, this is the 'essence' of those considerations and examples already given (*C*2, *C*3, *E*1, *E*2) and of those to come. The fact that, whereas, for any finite n, uniform partitions are allowed (all $p = 1/n$), in the countable case only extremely unbalanced partitions are allowed (under *C* and *N*), may serve as a 'symptom', which makes this 'restrictiveness' appear pathological.

We shall see, on the one hand, just how unbalanced they are, and, on the other hand, the objections to which this gives rise from a realistic point of view. The latter, of course, will vary according to the conception one holds.

*C*5 = *N*5 By taking the sum of probabilities to be = 1 (suppose we denote the probabilities by $p_1, p_2, \ldots, p_i, \ldots$, in decreasing order), one necessarily has an inequality such that for any $\varepsilon > 0$, however small, a finite number of events – the first n_ε – together have probability $>1 - \varepsilon$, and the infinity of the others together have probability $<\varepsilon$. (In such circumstances, I am tempted to say that the events 'are not countably infinite' but 'a finite number – up to trifles').

*E*5 The point made in *C*5 = *N*5 appears even more strange if we take as an example the following observation.

If, instead of the whole infinity of events, one only had the first $N = n/\varepsilon$ (where ε and $n = n_\varepsilon$ are as in the preceding case), there would be nothing to prevent one judging them equally probable (or almost so) in accordance with some assumed reasons or opinions. The total probability of the first n would then have been ε instead of $1 - \varepsilon$. Of course, even the infinity of probabilities could have all been taken $<1/N$, but the enormity of the inequality would reappear if we took some $n' = n'_\varepsilon$ and $N' = n'/\varepsilon$ to start with.

From a mathematical standpoint this is obvious What is strange is simply that a formal axiom, instead of being *neutral* with respect to the evaluations (or, for those who believe in them, with respect to the objective reasons), and only imposing formal conditions of coherence, on the contrary, imposes constraints of the above kind without even bothering about examining the possibility of there being a case against doing so.

3.11.6. Let us try to better imagine the reactions of individuals with different points of view.

*C*6 = *N*6 Suppose we are given a countable partition into events E_i, and let us put ourselves into the subjectivistic position. An individual wishes to evaluate the $p_i = \mathbf{P}(E_i)$; he is free to choose them as he pleases, except that, if he wants to be coherent, he must be careful not to inadvertently violate the conditions of coherence.

Someone tells him that in order to be coherent he can choose the p_i in any way he likes, so long as the sum = 1 (it is the same thing as in the finite case, anyway!).

The same thing?!!! You must be joking, the other will answer. In the finite case, this condition allowed me to choose the probabilities to be all equal, or slightly different, or very different; in short, I could express any opinion whatsoever. Here, on the other hand, the *content* of my judgments enter into the picture: I am allowed to express them only if they are unbalanced to the extent illustrated in *C*5–*N*5–*E*5. Otherwise, even if I think they are equally probable – as I would do in the case of *E*2 – I am obliged to pick 'at random' a convergent series, which, however I choose it, is in absolute contrast to what I think. If not, you call me *incoherent*! In leaving the finite domain, is it I who has ceased to understand anything, or is it you who has gone mad?

*C*7 = *N*7 In the same situation, an objectivist of the classical school finds himself facing case *E*2 (for him 'in conditions of symmetry all possible cases are equally probable').

This much is obvious: the infinite number of cases is equally probable and, therefore, they all have probability $1/\infty = 0$ (perhaps – he may think – I am not expressing myself in an orthodox fashion; the conclusion, however, is this one). To the objection of the teacher who wants a series with sum = 1, and who is not worried if one asks him whether he really wants an opinion so unbalanced as to give rise to the points raised in *E*5, he too will cry out: Is it I who has ceased to understand anything, or is it you who has gone mad?' And he will explain: 'I swear that I find myself in the ideal conditions of complete ignorance, with the absence of any reason to doubt whether any point has objective probability greater than that of any other one. In no other case can I be so sure of being able to state with precision that the objective probabilities are equal, because it is only in this case, where I cannot even see or distinguish the rational points, that I have reached the final sublime peak of total and unsurpassable ignorance. And now, what is the use of it? What are the objective probabilities I must give the various points, and how do I know which of them must be assigned a large probability, a small one, or a very small one?'.

*C*8 = *N*8 For the frequentist, this is even easier. If he thinks of a sequence of experiments (an ideal version of roulette, reduced to a point-ball which can stop at any rational point of the circle of *E*2) he will be in doubt as to whether a point will appear just a few times, or many times, or even infinitely many times. It is unlikely, however, that he will think for a moment that some point – and especially one which can be individuated right from the beginning – will appear so often as to have a limit-frequency different from zero.

*C*9 = *N*9 Here is a new and genuine mathematical objection to countable additivity: for those who conceive of probabilities as limit-frequencies (over a sequence, or, in von Mises' terminology, a 'Collective'), the fact that *limit-frequencies must satisfy finite additivity, but not countable additivity*, should be decisive.

(So far as I know, however, none of them has ever taken this observation into account, let alone disputed it; clearly it has been overlooked, although it seems to me I have repeated it on many occasions).

3.11.7. *C*10 A probability which is countably (but not perfectly) additive cannot be defined on the power set of the infinite set of events under consideration.

Therefore, it is necessary:

a) either to introduce restrictions that only allow one to refer to events given by certain 'subsets', excluding the others (in this case the logical justifications are not obvious, and the mathematical ones, which require the creation of special events by endowing the 'space' with topological properties, seem merely to have the status of 'Adhockeries for mathematical convenience');

b) or to accept perfect additivity, that is N, which appears *more logical* than C, for this reason in addition to that already given in $C1$ (but one encounters $N2$, and abandons any treatment in the continuum, even by means of the measure-theoretic model which is the actual aim of C);
c) or to accept finite additivity; that is A.

3.11.8. Do there exist objections to A (besides $A1$, which we have examined already)? In all honesty – and I shall willingly change my mind if any contrary evidence is brought to my attention – it seems to me that one should in general refer to prejudices and habits, rather than to objections. Independently of the discussion of specific aspects of the real problem (which are always neglected), it is these habits and prejudices which lead one to consider as 'natural', or 'absurd', those things in other branches of mathematics that are more or less customary, more or less up-to-date, and, above all, more or less 'convenient'. We refer to those fields where, in the absence of an intrinsic meaning, already existing and imposed from the outside onto the possible translations into mathematical definitions and axioms, it is admissible to choose those concepts and hypotheses that are most convenient, to choose them 'for mathematical convenience'.

We shall see something of these aspects and attitudes in Chapter 6 and in the Appendix. (It is often difficult to analyse them because they are more psychological than mathematical in character, and because one usually has to deduce things from odd comments rather than from explicit and systematic explanations.) If one wants to pick out an example of a sufficiently concrete position, having some validity,[34] I merely point to the following.

A2 It seems to many people that a countable partition that is not unbalanced (i.e. not reducing to cases 'finite up to trifles', as we jokingly called them in $C5$) is 'not feasible'. A positive integer N, unknown (random) and capable of taking on any value (between 0 and ∞, which is excluded), is always, in any practically or conceptually imaginable example, almost certainly not too large (and an upper bound is not given solely in order to avoid a more or less arbitrary choice). A partition of a set whose cardinality is that of the continuum, for example an interval, into a countably infinite number of (L-) measurable sets, is necessarily such that all the measure (except an arbitrarily small residual) is given by a finite number of them. They can be overlapping (as in the Vitali case) but then they are not measurable and, therefore, not even 'mentionable', and not even susceptible of a constructive description independently of the axiom of choice.

It is necessary to reply to this from various viewpoints.

A2a From the subjectivistic point of view – since, subject to the conditions of coherence, one has complete freedom of choice in evaluating the probabilities – one can perfectly well assign greater probability to a set with only one point than to a set which has very large measure, or is non-measurable. Conversely, can this line of argument justify attributing large probability to sets consisting of a single point and with small measure, and negligible probability to the large sets, leaving out the intermediate cases?

34 I hope that the reader can himself demolish the frequent attempts to 'prove' countable additivity under the tacit assumption of the validity of some property equivalent to it.

*A*2*b* Do not these examples themselves (although in a slightly more sophisticated manner) reveal the prejudice of assuming the measure-theoretic model as the universal one?

*A*3 Another plausible objection: all these examples and counterexamples are artificial, with no practical interest; there is no reason to prefer a less convenient theory simply because it allows us to take account of them.

*A*3*a* The examples have a critical function; to test the logical consistency of the various points of view. To accept the point of view which (I hope) they reveal to be the logically correct one does not imply that one has to occupy oneself with matters of this nature,[35] but only to avoid expressing oneself in a way that appears to be incorrect (albeit with reference to 'pathological' examples).

*A*3*b* Indeed, in practice, it will probably turn out to be advisable to limit oneself to *even simpler* ideas, sticking to the more elementary ambit (Jordan–Peano measure, Riemann integral) where the conclusions are unexceptionable, rather than passing to the more 'modern' set-up (Borel or Lebesgue measure, Lebesgue integral), given that the usual extension is based on a convention which is inadmissible as a general axiom, and difficult to justify in a realistic way as a particular hypothesis for individual practical cases. It seems to me that it is difficult to justify not only its validity, but even that possible interpretations and applications to actual and practical problems are not illusory.

*A*3*c* If we are going to talk about which theory is 'less convenient', we must distinguish the sense in which 'convenient' is to be understood. The theory given by *C* is, in general, more convenient to handle, and is convenient because it provides a well-determined answer in many cases where *A* just gives bounds. From the standpoint of *A*, it is wrong to substitute an exact answer in place of these bounds (and, anyway, inconvenient, since it forces us to exclude all those examples that might appear artificial, but which are not absurd). From some points of view, *A* is even more tractable; for example every limit of a probability distribution is, in *A*, a probability distribution (possibly not proper): this is not true in *C*. It is, in any case, a question of things which are logically relevant, not one of mathematical convenience.

*A*4 One more objection (a little premature as far as the applications it refers to are concerned, but not in terms of its formal meaning, nor for the understanding of example *E*6 below).

Proofs made in the spirit of *A* in order to invalidate the interpretations of *asymptotic results* (not yet discussed) as *limit-results* (deduced in accordance with concept *C*) often make use of the device of introducing a number N, which is 'chosen at random' (zero probability for each single n and finite segment $N \leqslant n$), assuming that from N onwards a certain process proceeds in a different way from that foreseen in the scheme of description.

This said, the objection is: *That's a different story: if the scheme changes, if there is a violent change, then the conclusions established under the assumption that the scheme remains unaltered, without foreseeing any possibility of a violent change, will certainly break down.*

35 Let us recall that the critical examples which Peano inserted into Genocchi's lecture notes, in order to show that certain 'theorems' did not always hold in 'pathological' cases, met with an exactly similar attitude of disapproval and incomprehension.

A4a Statements of this kind do not take account of the situation. The 'scheme', as usually described, does not explicitly foresee the possibility of a violent change, but it does not exclude it either: it is entirely neutral. It is, therefore, improper to refer to a 'violent change': the question of a violent change arises only when one adds to the mathematical scheme something more in the way of interpretation, which would be difficult to express. Indeed, if it were expressed, it would render trivial the result, which is beautiful and true only if one assumes that countable additivity is less restrictive than would appear from the following kind of example.

E6 As in *E2*, we can imagine 'choosing at random' a rational number in [0, 1] with a finite number of decimal places (all with the same probability (II!)),[36] the number of places being itself random, and not preassigned. If we think of a selection of the successive decimals (or of their successive deciphering or calculation, if they have been 'drawn all at once' and can be worked out successively, as for π), the process is clearly identical to that of drawing any real number whatsoever. At each drawing, all 10 figures have the same probability $\frac{1}{10}$, whatever the previous results may have been.[37]

If by 'catastrophe' we mean the exceeding of the last nonzero figure, it is certain that sooner or later this will happen. *But it will not be a catastrophe*: we will not be able to realize it; nothing will change in the described scheme. Even after 100 or 1 000 000 or 10^{1000} consecutive zeroes, provided we have no gift of divination, the probability that the next figure will be zero is $\frac{1}{10}$, as for any other figure; the probability that the next 100 figures will all be zero is 10^{-100}, as for any other 100-figure number; the probability that the figures will continue to be zero for evermore is zero, exactly as it is at any other instant, and after any arbitrary sequence of figures.

In this example, all the probabilistic assumptions explicitly stated for the process hold exactly; these lead to the conclusion that, with probability = 1, the 10 figures will each show up with limit-frequency $\frac{1}{10}$ (whereas, the limit-frequency is here = 1 for the figure 0, and = 0 for the others). The only assumption that does not hold is that of countable additivity, but if anyone considers it as an axiom, instead of a particular restriction (not valid in our example), he has the right (?) to omit its explicit statement and to check whether it holds.

3.11.9. *Conclusion* (*for the time being*). I do not know whether, and to what extent, the arguments put forward here have been persuasive. On the other hand, it is premature to accept or reject them before encountering other aspects of them and having seen their implications (in Section 3.12 following, at the end of Chapter 4, and in Chapter 6 and elsewhere, more or less incidentally). In view of this, however, I would like to have succeeded in convincing the reader of one thing; that we are dealing with a complex of

36 If one wishes, instead of choosing from this set one can imagine the choice of any rational whatsoever, as in *E2*. The rationals can be put together in 'equivalence classes' (where two numbers differ by a bounded decimal fraction; i.e. they coincide from some point on) and in each class *an identifiable representative can be chosen*; the one which is *periodic right from the beginning*. Every rational uniquely determines the components $r = p + d$ (p periodic, d decimal), and the sets I_d (of the r with the same d) give rise to a partition of the rationals into a countable number of sets superposable by translations (mod 1). To choose r is therefore a way of choosing d.
The partition is similar to that of Vitali for the reals, but here, fortunately, an infinite number of choices is not required.

37 We should refer to stochastic independence, but we shall come to it in the next chapter, Chapter 4, and here content ourselves with just mentioning the idea.

problems, connected and meaningful, concerning which there are many things to be discussed under various headings: the conceptual, the mathematical, the practical. It is not just, as might seem logical at first sight, a question of arbitrary conventions for the subtleties involved, having no connection with real problems.

3.12 Random Quantities with an Infinite Number of Possible Values

3.12.1. The above considerations obviously also apply to the case in which there are an infinite number of possible values for a random quantity X. Some new features also arise, however. We shall not concern ourselves with the general case until Chapter 6, but in the meantime it is necessary to mention certain refinements, although only for the more elementary case (elementary in a certain sense, at least) of a countable infinity of possible values x_h ($h = 1, 2, \ldots$). To these will correspond – or rather can be attributed by the person who evaluates them – probabilities p_h, either positive or zero (they might even all be zero), with

$$\sum_h p_h = 1 - p^* \leqslant 1, \left(0 \leqslant p^* \leqslant 1\right).$$

For any interval or set I, one could say, knowing only the x_h and p_h, that $\mathbf{P}(X \in I) = \sum_h p_h(x_h \in I)$ if the set contains a finite number of points, but only that

$$\sum_h p_h\left(x_h \in I\right) \leqslant \mathbf{P}\left(X \in I\right) \leqslant \sum_h p_h\left(x_h \in I\right) + p^*$$

if it contains an infinite number (given that the probability p^* can always be imagined as deriving solely from these).

3.12.2. In particular, if x is an accumulation point of the x_h (it does not matter whether it is one of them or not), we can have nonzero *adherent* probabilities, the latter defined to be the limit of $\mathbf{P}(x - \varepsilon < X < x)$ or $\mathbf{P}(x < X < x + \varepsilon)$ as $\varepsilon \to 0$ ($\varepsilon > 0$), and their sum (if we wish to distinguish, we refer to adherent from the left, adherent from the right). The adherent probabilities (or masses) cannot exceed p^*; not even if we take them all together, or even include those possibly adherent (from the left) to $+\infty$ and (from the right) to $-\infty$.[38] The adherent probabilities could not only have total probability $<p^*$ but also zero (in other words, nonexistent), although p^* was positive, or even $p^* = 1$. As an

38 One can either allow $+\infty$ and $-\infty$ to also appear among the possible values, or one can exclude them. Including them would entail thinking of X as a random point on the completed real line (compactified) with the adjunction of the 'extremes' $+\infty$ and $-\infty$. There is nothing absurd about this, although it is not usual to do and there is no point in insisting upon it. Every now and again we will make brief mention of such eventualities, but without entering into any obligation to observe case by case whether what is said is valid there also.

On the other hand, we must note a certain conflict of interest. As far as prevision is concerned (and here the inequalities are essential), the values $+\infty$ and $-\infty$ are distinct and very far apart (in fact, opposite). From an analytic point of view, however, it would be more natural to consider them as a single value (except for looking at it in terms of approaching from the left and right), thinking, for instance, of the complex sphere (and, in that context, of the circle of real numbers) and of functions which are 'continuous' there, like $y = 1/x$ at $x = 0$ (see *Matematica logico-intuitiva*, 3rd edn, pp. 124–133).

example: X = rational between 0 and 1, with the probability of each subinterval equal to its length (the uniform distribution).

3.12.3. The argument concerning the prevision $\mathbf{P}(X)$ is new and specific to this case. It is unnecessary to note that whatever one says concerning $\mathbf{P}(X)$ holds for any $\mathbf{P}(\gamma(X))$, where $Y = \gamma(X)$ is any function of X, whose possible values are $y_h = \gamma(x_h)$ with probabilities p_h (except that, if one of these values corresponds to an infinite number of the x_h, its probability may be, if $p^* > 0$, greater than the sum of the p_h instead of being equal to it).

What does the knowledge of the possible values x_h and their probabilities p_h allow us to say concerning $\mathbf{P}(X)$? Or rather, expressing ourselves in terms of what the question means in a (subjective) probabilistic sense, what restrictions does the knowledge of the x_h and an existing evaluation of the p_h (which we wish to remain coherent) impose on us when it comes to evaluating the prevision of X?

It is convenient to begin with the case of a *bounded* random quantity X, and to consider directly the minimum and the maximum of the accumulation points, which we denote by x' and x''; we therefore have

$$-\infty < \inf X \leqslant x' \leqslant x'' \leqslant \sup X < +\infty.$$

Let us prove that if $p^* = 0$ (i.e. if $\sum_h p_h = 1$, as it is if countable additivity holds) we must have the unique result $\mathbf{P}(X) = \sum_h p_h x_h$, as in the finite case. Apart from this special case we can only say that

$$\sum_h p_h x_h + p^* x' \leqslant \mathbf{P}(X) \leqslant \sum_h p_h x_h + p^* x''.$$

Thus, if we are not in the above case, $p^* = 0$, $\mathbf{P}(X)$ turns out to be uniquely determined if and only if $x' = x''$; in other words, if the x_h have a unique accumulation point, hence a limit to which they converge.

Proof. For a given $\varepsilon > 0$, take N sufficiently large so that we have

$$\sum_h p_h (h \geqslant N) < \varepsilon,$$

and put $X = X_1 + X_2 + X_3$ with

$$\begin{aligned}
&X_1 = X = x_h \text{ if } h < N, \quad \text{and} \quad \text{otherwise} = 0,\\
&X_2 = X = x_h \text{ if } h \geq N \quad \text{and} \quad x_h < x' - \varepsilon \quad \text{or} \quad x_h > x'' + \varepsilon, \text{and otherwise} = 0,\\
&X_3 = X = x_h \text{ if } h \geq N \quad \text{and} \quad x' - \varepsilon \leq x_h \leq x'' + \varepsilon, \text{and otherwise} = 0.
\end{aligned}$$

We have

$$\begin{aligned}
\mathbf{P}(X_1) = \sum_h p_h x_h (h < N) \to \sum_h p_h x_h \left(x_h \text{ bounded!}\right);\\
\varepsilon \inf X \leqslant \mathbf{P}(X_2) \leqslant \varepsilon \sup X,
\end{aligned}$$

because there are at most a finite number of possible values between $\inf X$ and $x' - \varepsilon$, and the same for those between $x'' + \varepsilon$ and $\sup X$, and the total probability of those between them with $h \geqslant N$ is the sum of a finite number of the p_h for which the sum of the series is $<\varepsilon$. Finally, we have

$$p^*(x' - \varepsilon) \leqslant \mathbf{P}(X_3) \leqslant p^*(x'' + \varepsilon).$$

All this holds for every ε and hence, as $\varepsilon \to 0$, one obtains the given bounds.

Remark. It is most instructive and important to observe that these bounds *cannot be improved on*; in other words, it is actually admissible to evaluate $\mathbf{P}(X)$ by giving it any value whatsoever between the two end-points (inclusive). The p^* resulting from infinite zero probabilities (distributed on the possible x_h; it does not matter if these already have positive probabilities $p_h > 0$ or instead have $p_h = 0$) could well be considered as deriving from an infinite number of the x_h converging towards x', or towards x'', and in any intermediate way.

(In addition, one notes that the proof neither presupposes nor establishes countable additivity: it holds here – as it may hold elsewhere – by virtue of additional assumptions implicit in the definition of the particular case.)

3.12.4. We pass from the case of bounded X to that of X *unbounded*. The case of one-sided unboundedness must be considered separately, and we therefore begin with the case of X unbounded from above (obviously, the analysis holds also for the other case); the general case follows as a corollary.

We also suppose that with certainty $X \geqslant 0$ (i.e. $\inf X \geqslant 0$); in the general case it is sufficient to put $X = X_1 - X_2$, $X_1 = 0 \vee X$, $X_2 = |0 \wedge X|$, in order to reduce everything to random quantities which are certainly nonnegative.

Moreover – in order not to complicate the exposition by encountering anew the circumstances already seen in the finite case – we suppose that there do not exist finite accumulation points. We can, therefore, suppose the x_h to be increasing, and tending to $+\infty$ as h tends to infinity.[39]

Under these conditions, putting

$$P_n = \sum_{h=1}^{n} p_h,\ P = \lim P_n,\ p^* = 1 - P,\ S_n = \sum_{h=1}^{n} p_h x_h,$$
$$S = \lim S_n,$$

we have

$$P_n = \mathbf{P}(X \leqslant x_n),\ \ 1 - P_n = \mathbf{P}(X > x_n),$$

p^* = the mass adherent from the left at $+\infty$, or placed at $x = +\infty$, or some here, some there;

$$S_n = \mathbf{P}\{X(X \leqslant x_n)\},\ S_n + x_n(1 - P_n) = \mathbf{P}(X \wedge x_n)$$

(the previsions of X either 'amputated' or 'truncated' at x_n; i.e. replaced, if X exceeds x_n, either by 0 or by x_n, respectively).

Since each 'truncated X' is always $\leqslant X$, we necessarily have

39 It is clear that the conclusions of this special case are essentially valid in general if one considers that $X' \leqslant X \leqslant X''$, where we set X' = (the smallest integer $\leqslant X$), $X'' = X + 1$ (and the unit of measurement can be taken as small as we please); X' and X' are automatically of the type considered (but to pursue this would introduce things which we reserve for the treatment of the continuous case).

If $X_\infty = +\infty$ exists among the possible values, it is not necessary that the finite possible values be unbounded (and not even that they be infinite in number) in order for us to be in the unbounded case.

$$\mathbf{P}(X) \geqslant S_n + x_n(1-P_n) \text{ for some } n, \text{and hence}$$
$$\mathbf{P}(X) \geqslant S + x_n(1-P) = S + x_n p^* \text{ for some } n$$

(because, if we let n increase in S_n and P_n, while keeping x_n fixed, the expression increases, but less than it would if x_n also were allowed to vary, and tends to the given limit).

It necessarily follows straightaway from this that $\mathbf{P}(X) = \infty$ if $S = \infty$ (the series $\sum_h p_h x_h$ diverges), or if $p^* \neq 0$ (there exists a probability placed at, or adherent to, $+\infty$), or both.

In the opposite case, $p^* = 0$ and S finite (the series of the p_h having sum = 1, and the series of the $p_h x_h$ being convergent), admissible evaluations of $\mathbf{P}(X)$ are given by

$$\mathbf{P}(X) = S = \sum_{h=0}^{\infty} p_h x_h, \text{ or any greater value, including } +\infty.$$

This is proved *by continuity* (and in the next section – Section 3.13 – we briefly discuss that property of continuity which we shall make use of here).

First of all, we set $X'_n = X(X \leqslant x_n)$ (X amputated) with $p'_h = p_h$ for $h \leqslant n, p'_h = 0$ for $h > n$, and $p'_0 = \sum p_h (h > n) = \mathbf{P}(X'_n = 0)$; as n increases, all the $p'_h = \mathbf{P}(X'_n = h)$ tend to p_h, but $\mathbf{P}(X'_n) = S_n \to S$.

We then set $X''_n = X'_n + a_n(X > n)$, in other words, X''_n (like X'_n) coincides with X if the latter does not exceed x_n, but when it does we replace it with a_n instead of with 0; a_n denotes the first of the x_h for which $x_h p_0 \geqslant n$.[40] The value a_n already gives a contribution $\geqslant n$, hence we certainly have

$$\mathbf{P}(X''_n) \geqslant n \to \infty.$$

We repeat the conclusions in a schematic form:

$$\text{in the case} \begin{cases} p^* > 0 & \qquad \mathbf{P}(X) = +\infty; \\ p^* = 0 \begin{cases} S = +\infty & \mathbf{P}(X) = +\infty; \\ S < +\infty & S \leq \mathbf{P}(X) \leq +\infty; \end{cases} \end{cases}$$

3.12.5. If X is unbounded from above and below, $\mathbf{P}(X)$ is completely undetermined. This is obvious straightaway from the fact that we could always have '$\infty - \infty$'; one can obtain this more rigorously by a passage to the limit in the previous cases (suitably balancing the positive and negative terms).

However, one might consider as *special* the evaluation which consists in taking, both for the positive part $0 \vee X$ and for the negative part $0 \wedge X$, the minimum (in absolute value) admissible prevision – denoting it by $\hat{\mathbf{P}}$ – and setting in general

$$\hat{\mathbf{P}}(X) = \hat{\mathbf{P}}(0 \vee X) - \hat{\mathbf{P}}(|0 \wedge X|) \left(\text{or, briefly, } \hat{S} = S^+ + S^-\right).$$

40 The argument, with a simple modification, also holds in the case in which $p_h = 0$ for all possible x_h from a certain $h = N$ on, so that $p_0 = 0$. One could, for instance, let $p_0 = (\frac{1}{2})^n$ taking this probability away from one or more of the p_h (for instance, from p_1 if $p_1 = 0$, starting from that n for which $(\frac{1}{2})^n < p_1$).

'Special' is *not used in a general sense* but if, and so long as, one can consider that, in a given case, the unbounded X is a theoretical schematization substituted for simplicity in place of an actual X, which is in reality bounded, but whose bounds are very large and imprecisely known.

This asymptotic prevision (as we shall call it for this reason) turns out to be:

$$\hat{S} = S^{+} + S^{-} \begin{cases} \textit{finite}, \text{ if } S^{+} \text{ and } S^{-} \text{are;} \\ \textit{infinite}, \text{ if one of the components is} : +\infty \text{ if } S^{+} = +\infty; \\ \quad -\infty \text{ if } S^{-} = -\infty; \\ \textit{undefined}, \text{ if both are infinite.} \end{cases}$$

3.13 The Continuity Property

The property says (and we shall make this precise and prove it) that *coherence is preserved in a passage to the limit.* The property does not hold (without further conditions) when we impose countable additivity. This turns out to be very useful as a tool in proofs of admissibility like the ones just given above (Section 3.12.4).

Theorem. Let $\mathbf{P}_n(E)$ be the evaluations of (coherent) probabilities defined over the same field of events $\mathscr{E}$ (or over different fields of events having $\mathscr{E}$ in common), and put $\mathbf{P}(E) = \lim \mathbf{P}_n(E)$ when it exists (letting $\mathscr{E}' \subseteq \mathscr{E}$ be the set of the E for which the limit exists). In this field the $\mathbf{P}(E)$ itself constitutes a (coherent) evaluation of probability.

Remark. In place of the (more 'familiar') formulation above, it would be (mathematically) preferable to substitute that in which one speaks of the prevision of random quantities rather than the probability of events, and hence of linear spaces (with appropriate definitions and convergence) rather than 'fields'.

Proof. The conditions of coherence are expressed by linear equations (or inequalities) involving a finite number of elements (events, or random quantities); in the passage to the limit these are preserved.

Remark. In a more expressive formulation (and more precise, so long as one recalls that the meaning of 'convergence' is that given above): an evaluation of probability $\mathbf{P}$ adhering to a set $\mathscr{P}$ of coherent evaluations is coherent.