STICHTING
MATHEMATISCH CENTRUM
2e BOERHAAVESTRAAT 49
AMSTERDAM

SP 103

(S 377)

J. Hemelrijk

Back to Laplace definition.

(Statistics Neerlandica 22 (1968), 13-21)

Back to the Laplace definition

by J. HEMELRIJK *

1. Introduction

This paper contains a plea for the return to, fundamentally, the classical (LAPLACE) definition of probability, with the inherent circularity removed.** This is to be done by defining a random drawing as clearly and exactly as possible and by coupling the definition of probability to drawing one element at random from a finite set. The transition to infinite probability fields and continuous distributions is seen as a mathematical convenience, leading to the measure-theoretical model.

There is not much that is new in this way of introducing probability, except perhaps the idea of openly returning to the classical definition instead of using it surreptitiously as seems to be the custom nowadays in elementary introductions to statistics and probability. The contention to the author is that there is nothing to be ashamed of in using this approach and that it can be made as exact as is necessary and possible in any applied science. This can be achieved by clearly distinguishing between reality (material reality that is) and mathematics (the mathematical model used to analyse reality). The gain is in simplicity and in obtaining a close connection between theory and practice.

2. Reality and mathematical model

This paper is not concerned with mathematics in itself, but with statistics and probability as a branch of applied mathematics. The approach is typical for an applied science: for a certain part of reality one has to build a mathematical model in order to profit by the advantage which mathematics has over practice, its greater exactness and its powerful arsenal of methods and knowledge. There are two transitions involved: first from reality to the model, then, after having solved the mathematical problem within the model, back to reality.

The importance should be stressed of distinguishing clearly and consistently between reality and the model. Confusing the model with reality is a fault which is frequently made, sometimes with disastrous effect.

Although this is not the subject of this paper, the author cannot resist the temptation to give one famous example: the classical Greek paradoxes. In one of these it is proved that Achilles cannot pass the tortoise, in another that an arrow cannot fly. The proofs are based on a mathematical model, and have baffled mankind for something like two thousand years. However, the main condition which a mathematical

Report S 377 (SP 103) of the Statistical Department of the Mathematical Centre, Amsterdam.

* University of Amsterdam.

** The author is aware of the fact that the definition proper can be traced back to earlier authors than LAPLACE; but it is usually called "LAPLACE-definition".

model should fulfill is, that the principal points of the part of reality it represents should be represented faithfully in the model. If not, then the model is not suitable for the situation, it is a bad model, or it is badly handled. The principal point of a race between Achilles and a tortoise is, of course, that Achilles *does* pass the tortoise. So, if this is impossible in the model, then the model is a bad one and should not be used. The Greeks' faith in mathematics, however, was so great, that they never really arrived at this point of thought. They (or at least many of them) thought mathematics more real than reality and it is this confounding of the mathematical model and reality which gave rise to this paradox and to some of the others. We are not, here, concerned with pointing out what was wrong in their model; it is sufficient for our purpose that it must be wrong because it does not represent reality faithfully. Many more examples of this kind can be given, also from modern society; especially in codes of law one often meets formulations where reality and model are freely confounded, sometimes much to the disadvantage of suspects.

The reason why it is necessary to distinguish carefully between reality and its model is that they are so very different. On the one hand a model always implies simplification and therefore it can never implicitly be trusted to be adequate. There is a well-known and generally applied remedy for this shortcoming: always to verify the practical conclusions, drawn from the model, by means of observation or experiment. On the other hand the model has the big advantage of exactness; only within the model mathematics can be used. Without this advantage no mathematical models would be used at all. It follows from all this, that in applying mathematics one has to make use of two different languages (or: at least two): *practical language*, which is always somewhat imprecise and vague, and *mathematical language*. In order to build the model one usually starts out by describing the practical situation under consideration in practical language. Building up the model then implies a translation of the practical terms into mathematical ones. The system of translation used is not universal; apart from some fundamental rules it may be different every time one builds a model. Therefore the correspondence between practical and mathematical terms should be clearly stated every time and the same system of translation should be used in translating back from the model to reality.

The point we want to stress most, because it bears on the rest of this paper, is that all practical language is more or less imprecise and ambiguous. This we have to live with – and we do – but it should not keep us from trying to be as clear and precise as possible. However, it would be unreasonable to demand a very high degree of exactness in things expressed in practical language. On the other hand exactness should certainly be insisted upon in all mathematical formulations. Therefore the reader should be informed, at any point, of the language used at that moment. We shall try to conform to this principle in the rest of this paper.

3. The statistical domain; randomizers

(Practical language). Mathematical statistics is a part of applied mathematics. There is no real need to say more about that, for we are not aiming at giving a definition of mathematics. The point is to outline the domain of statistics, i.e. the part of reality where statistics can be fruitfully applied. This may roughly be done as follows: statistics (and probability) can be applied in situations where individual outcomes show an element of *unpredictability*.

This brings us straight to the point: statistics aims at predicting in situations where individual outcomes are – to a greater or smaller extent – unpredictable. We have to analyse this concept, unpredictability, as far as possible.

The best way to do this seems to be to look for the most perfect form of unpredictability we know, unpredictability in vitro so to speak, and this we find in games of chance. These are based on some sort of *randomizer*, e.g. a roulette or some other kind of lottery, and since the randomizer is fundamental for our definition of probability, we shall give it our close attention.

A randomizer is a machine; a mechanical or electronic gadget which, when activated, gives one of a known finite set of results, one of the numbers 0, 1, ..., $k-1$ say. If we want to indicate that there are k possible outcomes we will call it a k-randomizer. The main property of a proper randomizer is, that its individual outcomes are unpredictable in the highest possible degree. Somewhat more precisely this can be expressed as follows: all methods of prediction, with or without knowledge or use of past results, are equivalent; none of them has an advantage over the others; there is no systematic difference in the number of hits they score. Or: there is no gambling system against a randomizer; it is impossible to win systematically when playing fairly against it. This concept was already used by VON MISES, but he worked it out in a rather impracticable way. It can also be expressed as follows: if the randomizer is given any odds over the player, however small, then the player will lose systematically.

This abundant use of different expressions of one fundamental idea is a sign of weakness. It seems to be impossible to give a really satisfying close definition of a randomizer in words. The reason for this can be detected by scrutinising the above descriptions: all of them are of a negative character, even though some of them are syntactically positive. The essential property of a randomizer is a negative one: its unpredictability.

There is, however, one very simple and effective way of defining material things: exemplification. Thus the unit of mass is defined as the mass of a certain piece of metal kept in an institute in Sèvres. And so we may say: a good roulette is a good randomizer, and everybody will understand this. Also, it is not difficult to make one yourself. The easiest way is perhaps to use a die, of indifferent quality, and to throw it two times in succession. Distinguishing only "even" (and denoting this by 0) and "odd" (1), there are four possible outcomes: 00, 01, 10 and 11. The results 00 and 11 do not count; if you get one of these you have to throw twice again. If you get 01 or 10, omit the last digit. The whole process then has two possible outcomes: 0 and 1, and there you have a very good 2-randomizer. If sufficient precautions are taken in the method of throwing, which it is easy to realise, the individual results are indeed very highly unpredictable. Instead of a die one may use a coin or even a button. The bias, which may be present in one throw is eliminated by using a pair and only noting the order of two different results. The principle of this trick is that, although it may be difficult to attain directly a very high degree of unpredictability mechanically, it is much easier to design an experiment "without memory", where successive results do not influence one another.

At this point the objection is usually made, that a randomizer, however skilfully or ingeniously made, nevertheless will never be a perfect randomizer. This is, of course, true. But it is true of everything: nothing is perfect in this world. The question is not whether an instrument is perfect – because it never is – but whether it serves its purpose. And there the proof of the pudding is in the eating. The proof that randomizers can be made adequate for at least one of their purposes lies in the fact that the people of Monaco do not pay any taxes: the randomizers in Monte Carlo take care of that part of the economy of the princedom.

Another objection, equally justified and equally futile, is that even a perfect randomizer could never be proved to be so. This is true, because it is impossible to prove that something (in this case a winning gambling system) does *not* exist. It is e.g. impossible to prove that ghosts do not exist, or psychokinesis. But if it does exist, thàt can be proved. If a machine is not a randomizer this can be proved by finding two methods of prediction which are not equivalent. Every statistician will here recognize the train of thought of the testing of a null-hypothesis.

Thus if a randomizer is to be used for a specific purpose, e.g. in the design of an experiment, and if one wants to test the randomizer before using it in the experiment, then one should test it by means of a statistical test which is also specific for this purpose.

That is: a test which is especially sensitive for those deviations from randomness which tend to invalidate the experiment one has in mind. The fact that in general such a test will be less sensitive for other deviations from randomness is of no importance, as these do not invalidate the experiment. As a matter of fact, what one needs for specific purposes is not absolute randomness, but relative randomness, with respect to the specific situation. But this is a subject in itself which is outside the scope of this paper.

Let us now abstract from the imperfections of material randomizers and consider an ideal one. Whatever is true for an ideal one will then be approximately true for a good one.

The equivalence of all systems of prediction for a randomizer implies that none of its possible outcomes occurs systematically more often than another. For, if in a 10-randomizer 0 occurs systematically more often than 5, say, then the system of always predicting 0 would be better than always predicting 5. Thus the unpredictability leads *logically* to equal frequencies. This point is so important that it is worth while to enlarge on it. What do we mean if we say that 0 occurs systematically more often than 5? Not that this is true in every finite sequence, but that one can safely and repeatedly predict more zero's than fives in a, future, long sequence. The vagueness in this formulation is the vagueness of all practical formulations. It is also inherent in the formulation of the *experimental* law of large numbers. Nevertheless it is clear enough if one gives an example: any reasonably well-made and well-thrown dice will systematically more often show a number ≤ 4 than > 4. Perhaps the term "equal frequencies", used above, should be avoided in this context, because in practice frequencies are usually unequal. What is meant is, strictly speaking, the absence of systematic differences between the frequencies of the results of a randomizer. An exact formulation of all this is only possible within a mathematical model: the theoretical law of large numbers; but this is only a *model* for the experimental one. Our conclusion is that a randomizer *must* satisfy the *experimental* law of large numbers, otherwise it would not be a randomizer. Thus the law of large numbers may be used as one of the ways to test a randomizer experimentally. If a machine satisfies this law, this is not sufficient to guarantee that it is a randomizer (counter-example: your watch). Unpredictability is a much stronger requirement than equality of frequencies in the long run. The moral of this is twofold. First: unpredictability of individual results is fundamental in a randomizer, the experimental law of large numbers is a logical consequence and therefore secondary. Second: unpredictability of individual results can only exist together with (approximate) predictability of frequencies in the long run. *Complete unpredictability does not exist in repeatable experiments.* This makes statistics possible!

A number produced by a randomizer is often called a *random number*. This leads time and again to considerable confusion. For if a k-randomizer can produce the numbers $0, \ldots, k-1$, then these are all random numbers. One should never, strictly speaking, use the adjective "random" for one number, or for a finite sequence of numbers. Infinite sequences we need not be concerned about, because in practice everything is finite, including sequences. The proper term, to be used instead of "random" is: "obtained from a randomizer" or "randomly obtained". It is not the number or the sequence that is random but the machine that has produced them is a randomizer. If the adjective "random" is understood in this way, there is no objection to its use as an abbreviation. We use it in this sense.

It is not difficult to see that a sequence of n random numbers from a k-randomizer is one random vector from a n-dimensional randomizer with k^n possible outcomes, all vectors of n numbers, each from $\{0, \ldots, k-1\}$. It is also a k^n-randomizer, for the unpredictability of individual outcomes remains unimpaired. Also, deleting some of the possible outcomes of a randomizer and asking for a new number if one of these occurs, does not impair the unpredictability of the other ones. These two properties enable us to use a k-randomizer as a k'-randomizer for any positive integers $k > 1$ and k'.

The predictability, implicit in unpredictability of individual outcomes, is not confined to frequencies in sequences of random numbers. Arithmatic operations with random numbers also increase the predictability of results. For example, although a pair of random numbers is also a random pair, the sum of two random numbers is already much less unpredictable.

To conclude, we repeat that the fundamental idea of a randomizer is of a negative nature: complete unpredictability of individual outcomes. It does not seem possible to get rid of this negative character. If this is true, it is useless to try. There is also another reason why it is perhaps better not to try. This is, that the main characteristic of a *statistical situation*, i.e. a situation where statistics can be applied, is of the same nature: it contains unpredictable elements. We can describe the domain of statistics as those practical situations, where observations or experiments behave, completely or in part, as if a randomizer were at work. This will sometimes be explicitly true, e.g. in random sampling or in randomized experiments, sometimes implicitly and in a very complicated way. Thus, if one observes the last numbers of the mileage of cars parked along an arbitrary street (without garages and car-factories), one has a rather complicated 10-randomizer.

In other cases one may have a situation which can be represented by a model with a randomizer, even though the practical situation is of a deterministic nature. The most extreme case of this is the use of pseudorandom numbers, generated by a recursion. It may be remarked here, that the question whether the unpredictability of a statistical phenomenon arises from ignorance of the observer or from some more fundamental cause, is not relevant at all in our context. A situation which might be analysed by means of deterministic method by somebody who knows quite a lot about it, may be analysed statistically by a person who knows less about it, so much less that for him some things are unpredictable which are predictable for the other. Both analyses are valid, although one may expect the best informed person to get the farther reaching and more definite results. Like everything else unpredictability is a relative concept and the questions whether perhaps "in the end" everything in this world is determined by causes and whether something like "chance" really exists (whatever that may mean), do not enter the picture at all. That is why the use of the word "chance" has been avoided in this paper.

To apply statistics in situations without random elements is not, in the opinion of the author, a promising activity. This is why he proposes to replace the principle of indifference by the positive requirement of the presence (explicit of implicit, real or imaginary) of a randomizer.

4. Relative frequencies on finite sets

The content of this section is trivially simple, so we can be short. (Mathematical language) Let

$$\Omega = \{\omega_1, \omega_2, \ldots, \omega_N\} \tag{1}$$

be a finite set of elements and let A, B, ..., denote subsets of Ω.

Let $N(A)$, $N(B)$, ... denote the number of elements in these subsets. Then we define the *relative frequency* $f(A)$ of A on Ω by

$$f(A) = N(A)/N. \tag{2}$$

The following well-known properties then hold

$$0 \leq f(A) \leq 1 \qquad \text{for every } A \subset \Omega \tag{3}$$

$$f(\Omega) = 1 \tag{4}$$

$$f(A \cup B) = f(A) + f(B) \qquad \text{if} \quad A \cap B = \emptyset. \tag{5}$$

Conditional relative frequencies, under the condition C, where C is a non-empty subset of Ω, are defined as relative frequencies on C instead of Ω. Two subsets A and B are called *orthogonal* on Ω if

$$f(A \cap B) = f(A)\, f(B) \tag{6}$$

and, if B is non-empty, this is equivalent to

$$f(A|B) = f(A) \tag{7}$$

where $f(A|B)$ is the conditional relative frequency of A under condition B.

5. Probability

(Practical language) Consider the following situation. A set of N objects is given and a randomizer. By means of the randomizer we are going to choose one object from the set. Some of the objects have a property A, the others do not. Will the object, indicated by the randomizer, have property A?

In this situation we will say that there is a probability $P(A)$ of A occuring, with

$$P(A) = N(A)/N, \tag{8}$$

$N(A)$ being the number of objects having property A.

This clearly is the Laplace definition, but now only to be used if there is a randomizer at work.

(Translation into mathematical language) The mathematical model for the above statistical situation is a set Ω, of section 4. The element to be pointed out by the randomizer is indicated by $\underline{\omega}$.* The use of an underlined symbol will always indicate a randomizer at work.

(Mathematical language) To the notions given in section 4 an undefined new entity $\underline{\omega}$, called a *random element* of Ω, is added and, for every $A \subset \Omega$, another undefined notion $\underline{\omega} \in A$, called a (*possible*) *event* and we define

$$P(\underline{\omega} \in A) = f(A), \tag{9}$$

the probability of $\underline{\omega} \in A$.
The rules for events are the usual ones, e.g.

$$\underline{\omega} \in A \quad \text{and} \quad \underline{\omega} \in B \leftrightarrow \underline{\omega} \in A \cap B,$$

etc. They are indentical to the rules for ω not underlined.
From definition (9) and (3), (4) and (5) now follows:

$$0 \leq P(\underline{\omega} \in A) \leq 1 \qquad \text{for every} \quad A \subset \Omega \tag{3'}$$

$$P(\underline{\omega} \in \Omega) = 1 \tag{4'}$$

$$P(\underline{\omega} \in A \cup B) = P(\underline{\omega} \in A) + P(\underline{\omega} \in B) \qquad \text{if} \quad A \cap B = \emptyset. \tag{5'}$$

* Or another notation which distinguishes clearly between random and non-random.

The latter may equivalently be written:

$$P(\omega \in A \text{ or } \omega \in B) = P(\omega \in A) + P(\omega \in B) \quad \text{if} \quad A \cap B = \emptyset \qquad (5'')$$

and "if $A \cap B = \emptyset$" is often replaced by the expression "if $\omega \in A$ and $\omega \in B$ are mutually exclusive". The abbreviation $P(A)$ for $P(\omega \in A)$ is often used.

Further elaboration is straightforward. E.g. the events $\omega \in A$ and $\omega \in B$ will be called *independent* if, and only if, A and B are orthogonal, etc. The model, thus obtained, is called a finite *probability field.*

6. Comment

(Practical language) Thus we arrive painlessly at a finite model for probability theory. For practical purposes this should be sufficient – for in practice everything is finite – but if we stopped here we should cut off the main part of mathematics and this we do not want to do. So in a later stage we have to generalize our model. But some remarks about the finite model have to be made first.

According to the above definition of probability this term can only be used for *one* random drawing from a (finite) set Ω. Every more complicated problem has to be reduced to this situation, e.g. by using product-fields for independent random drawings. This is completely in accordance with the measure-theoretical way of introducing probability, with this difference only, that there there is no need to go back to equal probabilities for the elements of Ω. If Ω is infinite this is, moreover, impossible. Remaining, for the moment, in the finite domain, it is clear that all probabilities will have rational values and that, on the other hand, all rational probabilities can be constructed.* A general, rational and finite, probability field can be obtained from Ω by bunching together relevant elements of Ω, i.e. by dividing Ω into N' $(<N)$ distinct subsets and mapping these onto a new set Ω', preserving probabilities. This is a well-known method in measure theory too.

An objection, which has often been made to the Laplace definition is, that often it is impossible in practice to indicate the "equally probable" cases which are needed to compute the probabilities involved. The usual example is a throw with a biased die. This difficulty is not pertinent for our setup. The reason why we wish to use the notion of probability for the possible results of a biased die is, that its behaviour is equivalent to drawing an element at random from a set of, say, tickets, each of which carries one of the numbers 1, 2, ..., 6. The die being biased, the relative frequencies of the numbers in this set are unequal and, in most cases, unknown. But the model is clearly pertinent. If it were not, i.e. if a biased die behaved fundamentally different from such a lottery, then probability would not be a pertinent notion for that die.

An important point in applying statistics is the way back from the mathematical model to practice. The main point is, of course, the translation of a probability, e.g. a confidence coefficient, $1-\alpha$ say. This is usually done by means of the frequency

---

* One should not forget that real numbers do not "exist" in practice, where they are always broken off after a finite number of decimals, but only in mathematics.

interpretation but it seems more elegant to put that interpretation in the second place, where it belongs, and to use the fundamental idea of a randomizer. The practical explanation would then be as follows. If you use this method of analyzing your data, you will get an answer that may be right or wrong. It will not be possible to be certain whether it is right or wrong. Imagine that at the same time when you apply the method, a randomizer draws one element from a set of elements, consisting of a fraction α of "black" and $1-\alpha$ of "white" elements. If a "white" element is drawn the answer is right, otherwise it is wrong. But you are not allowed to look at the element drawn by the randomizer. This seems to describe the nature of a confidence coefficient very precisely. It also emphasizes strongly that the confidence coefficient refers to the method used and not to the individual result obtained. This is an important point, for one should never say, as is often done, that the result obtained has a probability $1-\alpha$ of being right; it is the method of obtaining the result which the probability $1-\alpha$ refers to.

Of course the frequency interpretation may be used as a help, because a randomizer obeys the law of large numbers. But this is only secondary and using the above approach probability can also be applied to unique situations, i.e. situations where repeatability is absent or practically irrelevant. Moreover, in this way a set-up is obtained which in principle allows the introduction of Bayesian and subjectivistic points of view, as far as these can be represented by (imaginary) randomizers. It seems to the author that this could certainly be done to a great extent, although he does not feel inclined to do so himself.

7. Generalisation of the model

It would be very awkward to stop at finite probability fields, for finite mathematics is only a small and awkward part of mathematics. But generalisation to the measure-theoretical model is now straightforward. The main feature is the generalization of property (5′) to an enumerable sequence of events and this, having no real analogue in practice – where, it can never be said often enough to mathematicians, everything is finite – should be done in the form of an axiom. So then it is perhaps better to introduce the whole model axiomatically. But the link between practice and mathematics is strengthened by first introducing probability in the way indicated here and it also has definite didactical advantages.

One consequence one can draw, for the measure-theoretical method, from the above considerations and which, I think, would be a definite improvement, is to introduce the random element ω of Ω into the measure-theoretical model at the moment when probability is introduced. This would be a clear landmark separating probability theory from measure theory proper.

Reference

HEMELRIJK, J. (1966), Underlining random variables, Statistica Neerlandica **20**, 1–7.