stichting
mathematisch
centrum

AFDELING MATHEMATISCHE STATISTIEK SW 53/78 APRIL
(DEPARTMENT OF MATHEMATICAL STATISTICS)

J. HEMELRIJK

RULES FOR BUILDING STATISTICAL MODELS

Preprint

2e boerhaavestraat 49 amsterdam

Printed at the Mathematical Centre, 49, 2e Boerhaavestraat, Amsterdam.

The Mathematical Centre, founded the 11-th of February 1946, is a non-profit institution aiming at the promotion of pure mathematics and its applications. It is sponsored by the Netherlands Government through the Netherlands Organization for the Advancement of Pure Research (Z.W.O).

AMS(MOS) subject classification scheme (1970): 60A05, 62A99

Rules for building statistical models *)

by

J. Hemelrijk

SUMMARY

One of the fundamental questions in statistical model building is when to use the same model for different situations or experiments. Axioms, however useful mathematically, say nothing about this. The author therefore proposes to introduce rules for the choice of a statistical model which have the character of instructions for use of the statistical toolkit. One of the basis rules proposed is the "principle of equivalence". Two repeatable experiments are called statistically equivalent if they cannot be distinguished from one another by means of sequences of outcomes of arbitrary length. This principle is elaborated and illustrated by means of an example. If experiments are (deemed to be) statistically equivalent the use of the same statistical model for all of them is justified. This principle is then used for the introduction of conditional probabilities and composite models, with symmetric probability spaces as models for randomizers as a starting-point.

KEY WORDS & PHRASES: *foundations of statistics, statistical models, statistical equivalence*

---

*) This report will be submitted for publication elsewhere

CONTENTS

"Statistics uses the empirical hypothesis that apparatus ('lotteries') exist, admitting random choices of one among any given number of elements. Such apparatus do not exist in absolute perfection and their degree of perfection can only be defined after development of their theory. Their role is analogous to that of rigid bodies in euclidean geometry and of perfect clocks in dynamics. Empirical interpretation of probability statements is only possible with reference to such random apparatus or to natural phenomena empirically found to behave statistically sufficiently like these".

D. van Dantzig (1957)

1. INTRODUCTION

The use of *mathematical models* is widespread and of an old date, but the general recognition of this fact is comparatively new. The question of how to choose a statistical mathematical model has led to considerable confusion and controversy, and still does. Mathematical statisticians wisely save their skins by using the axiomatic approach, leaving the controversy to others and the confusion to the users of their theory. For axioms, however useful, say nothing about their application. It seems to the author that the time has come to formulate *rules* for the choice of statistical models. In this paper a number of such rules are proposed. They will certainly not please everybody, if only because they are formulated from the classical, objectivistic, point of view. They may, however, strengthen and clarify this point of view and help the user of statistical theory in its correct use and interpretation.

The subject is an extensive one, which can only be touched upon in a short paper. Therefore many details have to be taken for granted, the history of the subject is left aside and the controversy between objectivists and subjectivists is ignored.

In general a mathematical model is a *simplification* and an *exactification* of a part of reality. The simplification is necessary because of the extreme complexity of reality and the exactification because of its vagueness. Reality is always a bit out of focus: equality, for instance, is usually approximate equality and therefore not strictly transitive. In a mathematical model transitivity of equality and other desirable properties hold exactly and this makes it possible to develop extensive theories. But one should keep reality and model strictly apart. Confusing the two leads to baffling paradoxes - some of them well known - which can only be solved by disentangling reality and model.

Statistical models are concerned with parts of reality which are subject to uncertainty and which we will call (statistical) experiments. The possible outcomes of a statistical experiment are usually known, but the actual results are in a higher or lesser degree unpredictable. Causality does not seem adequate for analysing such experiments; instead the probabilistic approach is used.

In the following sections rules for using this approach are formulated step by step. These rules are not part of mathematics. They are not theorems nor are they laws of nature. They may be seen as *directions for use* of statistical models. They are certainly not perfect (nothing is) and their use cannot be enforced. But they are useful as a guide for sensible application of statistical methods.

2. RANDOMIZERS

Pure unpredictability in a statistical sense is found in a lottery, or randomizer. Everybody knows what a lottery is, but nevertheless it is suprisingly difficult to give a satisfactory description of its properties. A separate paper would be needed to this end. Let us just point out some

the probability of an event is equal to the ratio of the number of possible outcomes favorable for the event to the total number of possible outcomes.

More precisely, if we call the activation of an N-randomizer: "*drawing at random from* 0,...,N-1", then the model for one random drawing consist of three elements:

1) *The space of (elementary) events:* $\Omega = \{0,\dots,N-1\}$.*)
2) *Composite events: all subsets of* Ω.
3) *The Laplace-definition assigning a probability to every event:*

(1) $P(\Omega') \overset{\text{def}}{=} N(\Omega')/N \qquad \Omega' \subset \Omega$

with $N(\Omega')$ = the number of element of Ω'.

The basic threefold structure of this model holds for all statistical models, though usually in a more complicated form. It is also completely in harmony with the axiomatic set-up. We call this model a *finite symmetric probability space* and our first rule is:

Rule 1. *For one random drawing we use a finite symmetric probability space as mathematical model.*

Now consider a sequence of n random drawings, resulting in an n-vector of numbers from Ω. According to property c) of section 2 this composite experiment is the same as one random drawing from the N^n possible n-vectors. Thus rule 1 also gives us the model for this sequence of drawings. If one works this out the result is the product probability space of n finite symmetric probability spaces, one for each of the n random drawings. We omit the details; they are well-known to every statistician and we want to hurry on to more important points. But we do remark that the reasoning also holds for a sequence of random drawings from different randomizers and that we arrive thus at our second rule:

Rule 2. *For a sequence of* n *random drawings we use as a model the product probability space of the* n *symmetric probability spaces of the separate drawings.*

Remark that the term "independent" need not yet be introduced at this stage; it is implicit in property b) and emerges explicitly in a natural way when later on conditional probabilities are introduced. At the present stage one

---

*) "(possible) result", "(possible) outcome" and "elementary event" are used as synomyms.

might say that a randomizer is independent of everything: it walks, like a cat, by itself.

4. THE PRINCIPLE OF EQUIVALENCE

The transition from rule 1 to rule 2 has been accomplished by stating that - according to property b) - n random drawings from $0,\ldots,N-1$ "are the same as" one random drawing from $0,\ldots,N^n-1$ (after numbering the n-vectors in an arbitrary order). This expression "the same as" is not very accurate; the two experiments compared are not the same, but they both have the properties of a randomizer. In a certain sense they are equivalent with respect to their statistical properties. It is worth while to elaborate on this point because it leads us to one of the key-points of our set-up.

Consider two repeatable experiments E' and E" with the same possible outcomes ($\Omega' = \Omega''$) but otherwise possibly very different. Let the following information be supplied:

1) *an accurate description of* E' *and* E",
2) *two sequences of results* A *and* B *from these experiments, however without identification; this means that it is not known whether* A *and* E' *(and* B *and* E"*) belong together or the other way around.*

Additional information is supplied on request:

3) *further details about* E' *and* E",
4) *extensions of the sequences* A *and* B *(again without identification),*
5) *sequences* C' *from* E' *and* C" *from* E".

If in this situation there is no conceivable method of identifying the sequences A and B, then E' and E" are called (statistically) *equivalent*. Their statistical behaviour with respect to the possible outcomes considered, is the same. The generalization to more than two experiments is straightforward and we can now formulate:

The principle of equivalence. If experiments are equivalent in the sense described above, then the use of the same model for all of them is justified.

External reasons like practical importance and cost of time and money may lead to the use of different models when, statistically speaking, the use of the same model would be desirable. In this paper, however, we will strictly adhere to the principle of using the same model for equivalent experiments.

A lot more can be said about the concept of equivalence, but a practical example may at this point be more clarifying. Five experiments have, to this end, each been excecuted 221 times. They are

E_1: *recording the last digit of the hodometer**) *of the authors car when he left the car for more than half an hour.*

E_2: *recording at the same moments, the last digit of the sub-hodometer, which records the same distance in units of* 100 *m (mod* 10 000*).*

E_3: *recording the last digit of the hodometers of cars in public parking lots.*

E_4: *throwing a blue tensided die carrying the numbers* 0,...,9.

E_5: *throwing a red tensided die with the same numbers.*

For every experiment the results of 221 excecutions were recorded in the order of their observation. The dice were well made and they were thrown in such a way that E_4 and E_5 may be considered to be 10-randomizers. For these two experiments equivalence is clear: from property b) of section 2 it follows that all N-randomizers are equivalent (for any fixed N). It is not very plausible that E_1 and E_2 are equivalent to E_4 and E_5, but E_3 might well be. For although E_3 is much more complicated then E_4 and E_5 it is difficult to imagine why it would be possible to find two systems of predicting the next outcome of E_3 one of which is better than the other. This might well be possible for E_1 and E_2.

It is clear that speculations of this kind are not a sufficient basis for deciding about equivalence. The observations themselves, however, may help. And one of the tasks of statistical theory is to provide methods to test equivalence of experiments and the goodness of fit of models to experiments. These methods are indeed available and one of them can be used in our case. In order to confuse the reader the five sequences have been assigned labels A, B, C, D, E at random. Table 1 contains the observations in their original form. It is difficult to draw any conclusions directly from these date. They have been completely recorded in Table 1 in order to enable the reader to play around with them himself. A first step in getting a better survey of the data is to arrange them in a frequency table.

---

*) The hodometer cumulatively counts the distance covered by the car in km (mod 100 000).

This has been done in table 2, where two columns have been added, one for the well-known test statistic χ^2 and one for the right-hand tail-probability P. Extreme values of P indicate deviations from randomness; the number of degrees of freedom is 9, the hypothesis tested: randomness.

Table 1. Five sequences of observations

A	9 0 4 9 3 3 6 1 8 7 7 5 9 5 3 8 1 8 5 3 9 9 7 9 2 3 5 2 7 0 0 6 4 6 5 5 1 9 4 5 4 7 0 4 2 5 1 7 8 5 6 3 1 4 6 1 3 3 4 7 3 1 9 3 8 9 8 0 6 0 6 5 6 6 4 3 9 0 0 3 8 9 9 0 8 5 3 4 2 5 5 4 7 2 5 2 0 8 7 3 7 0 6 9 6 4 9 1 3 2 3 9 7 0 3 2 6 6 2 8 9 6 8 8 2 8 1 0 8 5 3 3 8 8 0 5 7 7 2 4 7 1 0 6 6 0 5 5 6 6 4 4 9 9 0 7 8 3 7 4 5 2 3 1 3 7 9 4 9 8 4 0 4 6 4 1 3 7 9 3 3 0 8 4 3 8 7 4 4 4 7 9 3 8 8 7 9 1 2 7 4 6 1 8 2 7 7 8 8 7 8 3 3 4 7 8 9 3 2 2 7
B	7 1 2 8 8 1 8 9 5 4 0 6 0 0 0 1 1 6 7 4 7 5 2 8 3 1 3 5 4 6 7 7 2 9 4 9 9 9 1 2 4 7 3 1 6 4 9 7 6 3 8 8 3 6 8 9 6 9 5 9 3 8 5 1 7 9 1 7 3 6 8 1 8 8 2 1 3 0 2 3 0 4 5 0 7 8 0 5 2 9 3 4 3 4 5 1 8 6 1 3 8 5 3 2 0 1 3 9 6 7 9 2 4 6 4 0 4 6 3 9 5 7 8 4 1 6 1 2 2 7 3 9 4 2 6 5 1 9 6 2 8 7 2 8 6 4 5 5 0 7 0 0 4 9 5 1 5 8 7 9 4 8 3 9 3 2 3 2 6 6 8 5 1 5 2 9 8 2 7 7 5 2 0 2 0 6 1 9 5 1 7 4 0 1 7 4 4 2 1 4 0 6 3 6 0 9 8 2 2 5 1 1 8 5 1 6 0 8 7 9 8
C	4 5 7 9 0 2 7 3 7 3 0 1 1 0 9 5 5 4 0 9 2 3 4 1 9 8 8 8 6 0 6 3 5 0 6 3 6 4 3 0 9 6 4 0 1 6 1 5 0 2 3 8 9 7 6 8 7 8 9 6 2 6 0 2 5 5 8 8 3 9 3 2 2 0 5 2 1 2 7 1 0 1 7 8 9 2 0 1 3 8 2 1 1 3 6 1 5 0 4 4 8 3 5 6 7 5 5 9 8 5 7 7 2 3 8 8 1 7 8 0 0 0 0 7 8 4 9 9 9 5 0 1 6 7 1 5 8 2 9 2 3 6 4 9 6 5 5 5 9 1 9 2 6 5 6 1 7 9 7 3 0 4 0 8 2 4 3 2 3 5 7 0 7 0 7 8 3 0 6 7 2 7 1 9 4 9 9 9 9 4 1 9 7 6 9 2 1 5 8 3 7 1 0 2 0 6 6 4 1 2 1 0 0 9 0 6 2 7 1 5 1
D	5 6 6 0 8 2 6 5 3 9 2 6 5 5 4 5 4 8 2 9 5 6 7 7 2 9 9 8 7 3 0 3 5 9 2 9 3 2 2 4 1 5 9 6 2 8 5 1 1 4 7 9 3 3 4 6 4 8 4 7 4 0 4 8 1 5 0 3 7 7 3 1 4 3 6 5 9 3 2 3 9 8 4 8 1 1 4 3 9 2 3 4 3 9 5 9 2 9 6 4 5 9 5 7 3 8 3 6 0 6 2 3 4 5 4 1 6 0 4 7 1 7 3 7 1 2 4 4 4 7 1 5 0 1 3 2 5 9 1 7 0 4 8 7 9 8 3 3 7 0 2 8 7 6 5 9 2 5 1 0 1 9 3 6 5 6 8 8 4 9 2 8 1 9 3 6 5 1 4 2 6 0 3 3 2 7 7 0 4 8 1 5 2 2 5 9 8 2 4 6 1 0 4 7 5 7 5 3 3 7 1 0 5 9 3 6 9 9 8 7 6
E	7 3 9 5 9 5 1 3 4 7 0 0 0 9 5 1 9 6 5 9 0 1 9 6 9 4 4 1 6 3 7 0 0 4 9 0 7 8 6 7 3 3 1 0 8 7 3 2 5 1 9 7 9 1 0 4 7 9 6 2 1 4 2 2 2 9 9 3 5 9 6 2 1 8 1 0 1 9 4 1 0 3 7 6 3 4 3 1 8 8 9 8 8 8 0 5 1 6 5 1 3 8 0 5 8 6 8 2 5 7 6 5 8 2 9 3 0 9 7 5 5 9 3 1 9 4 2 4 4 3 5 5 3 8 9 7 0 1 4 6 2 5 3 4 7 0 1 0 5 9 4 4 4 6 3 0 0 9 4 9 1 3 6 4 8 4 5 8 2 2 3 5 8 8 2 6 8 3 5 9 7 3 8 4 7 0 3 4 7 9 2 3 7 3 3 8 8 7 9 9 2 5 4 1 9 2 1 4 9 2 1 6 0 3 4 6 6 5 3 4 0

Table 2. Frequencies of 0,...,9 in the five sequences

	0	1	2	3	4	5	6	7	8	9	χ^2	P
A	19	14	16	30	25	19	20	27	27	24	11.26	0.26
B	19	26	23	20	21	21	22	21	25	23	1.94	0.992
C	29	25	22	19	14	22	21	23	20	26	6.92	0.65
D	14	20	22	28	26	26	20	22	18	25	7.46	0.59
E	22	22	17	27	25	22	17	18	21	30	7.46	0.59

None of the frequencies in table 2 deviates extremely from its mean 22.1. None of the values P is very small. One, however, pertaining to sequence B, is very close to 1, indicating some source of regularity which cannot be expected in a randomizer. Thus B may well stem from E_1 or E_2. But the result is still very undecisive. Therefore we go one more step in our analysis, aiming straightly at a point where E_1 and E_2 may well be very different from E_3, E_4 and E_5. For every pair of consecutive results, x_1 and x_2 say, we form the difference $x_2 - x_1$ (mod 10). This gives us five new sequences of 220 results each. We need not give a table of these in the form of table 1, because the reader can easily write this down himself. The new sequence A would start with: 1 4 5 4 0 ... This operation applied to successive results of a 10-randomizer gives again a 10-randomizer. This can easily be proved by means of the model implied by rules 1 and 2. It can also be viewed as a property like c) and d) of section 2; the reader can easily verify this by some thinking. On the other hand it is very plausible that this does not hold at all for E_1 and E_2 because the author often travels the distance from home to work by car.

Table 3. Frequencies of 0,...,9 in differencies mod 10

	0	1	2	3	4	5	6	7	8	9	χ^2	P
A	28	26	22	21	18	24	20	29	24	18	4.82	0.89
B	17	20	23	12	19	28	37	28	18	18	21.27	0.012
C	24	29	18	16	25	28	22	18	16	24	9.36	0.40
D	18	15	14	35	49	10	26	11	11	31	68.64	2.6×10^{-9}
E	23	19	22	31	22	20	19	16	23	25	6.82	0.66

The frequencies of the five new series are given in table 3. Now the situation is completely changed. In D the differences 3 and 4 are very predominant and in B the same holds, but less strongly, for 5, 6 and 7. The value of P is very small for D and small for B; there is little doubt that D and B stem from E_1 and E_2, possibly even in this order. Additional information of the types 4) and 5) mentioned above would most probably lead to a decision in this question. Thus our conclusion is that E_4, E_5 and E_6 may well be considered equivalent, but E_1 and E_2 certainly are not equivalent, neither to each other nor to the other three. If the reader would wish to try to identify E_3 among A, C and E, he can provide additional observations of E_3 himself.

Anticipating an objection to the principle of equivalence we may concede that it will never be possible to prove conclusively that two experiments are equivalent. But then, absolute certainly about such things is not part of this life. If experiments are *deemed* equivalent for sufficiently sensible reasons and if observations in sufficient numbers do not contradict this, then the principle can be used. For on the other hand non-equivalence can be proved experimentally to a reasonable degree of certainty, as the example illustrates.

5. PROBABILITY SPACES WITH UNEQUAL PROBABILITIES

To arrive at probability spaces with *unequal probabilities* for the elementary events, the space of events Ω of a symmetric probability space is partitioned into a set of non-overlapping subsets. These, together with their probabilities form a new probability space. The *addition law for exclusive events*, which in the symmetric probability space follows from the Laplace-definition, is carried over to the new probability space and this leads us to finite discrete probability spaces. The principle of equivalence then justifies the use of such a space as a model for experiments where a lack of symmetry does not suggest the use of equal probabilities at all. A simple example: let experiment E' be throwing a loaded six-sided die, E" using an N-randomizer with sufficiently big N with $\Omega = \{0,\dots,N-1\}$ partitioned into six subsets with unequal numbers of elements $n_1,\dots,n_6$, carrying the numbers $1,\dots,6$. The contention is that for suitably chosen N and $n_1,\dots,n_6$ the two experiments are equivalent, thus justifying the use of a discrete probability field for E'. Of course a suitable choice of $n_1,\ldots,n_6$ (and N) would have to depend on observations of E' but that only emphasizes the need of testing the goodness of fit of a model which has been chosen on the basis of rules and practical considerations. We thus arrive at:

Rule 3. *If an experiment is equivalent to a suitably chosen partitioned randomizer then we use a discrete probability space as mathematical model.*

6. STATISTICAL INDEPENDENCE

At the end of section 3 it was remarked that the independence of successive uses of randomizers is implicit in the properties of a randomizer. It is expressed in property b) by means of the fact that the past does not help to predict the future. This concept must be generalized and more formally expressed:

DEFINITION. *Consider* n *experiments* $E_1,\ldots,E_n$, *each of which separately is adequately described by a completely specified probability space; if knowledge of the results of any part of these experiments (after they have been performed) does not influence the predictability of the results of any of the others, then the experiments are called statistically independent.*

This, again, is a practical concept of considerable vagueness, which needs exactification by means of a mathematical model. It is clear from the definition and the previously formulated rules that the whole sequence $(E_1,\ldots,E_n)$ is equivalent to n random drawings from suitably chosen partitioned randomizers and thus rule 2 indicates the use of the product-space:

Rule 4. *If* n *statistically independent experiments are each described by a probability space the combined experiment* $(E_1,\ldots,E_n)$ *is described by the product of these probability spaces.*

Omitting, in this rule, the term "completely specified", which figures in the above definition, only means a slight generalization. The term cannot be omitted from the definition: if there are unknown parameters involved previous experiments - independent or not - may supply information about these parameters and thus influence the predictability of the other experiments. This would for instance occur in a sequence of throws of the loaded die used as an example in section 5, where nevertheless successive throws would be independent.

7. CONDITIONAL PROBABILITIES AND COMPOSITE MODELS

From independence to dependence is only one step but a very important one. An example of statistical dependence is found in repetitions of E_1 of section 4; line D of table 3 (which does in fact pertain to E_1) clearly indicates that adding 4 to the previous result (mod 10) is certainly superior as a method of prediction to adding 5. To build models for dependent experiments we need conditional probabilities.

Usually conditional probabilities are introduced in the model by means of a definition. Let Ω_1 and Ω_2 be subsets of the space of events Ω then the *conditional probability* of finding an element of Ω_2 under the condition that an element of Ω_1 occurs is

$$(2) \qquad P(\Omega_2 | \Omega_1) \overset{\text{def}}{=} P(\Omega_1 \cap \Omega_2)/P(\Omega_1),$$

where $P(\Omega_1)$ must be positive. This definition in itself says nothing about the way it should be used in applications. We therefore present a justification of (2) based on our rules, which also leads to a new rule giving insight in the way it should be used for model-building.

Consider the following two experiments.

E': *drawing one element at random from* Ω_1 *(using an* $N(\Omega_1)$*-randomizer for the purpose),*

E": *drawing elements at random from* Ω *(by means of an* $N(\Omega)$*-randomizer) until for the first time an element from* Ω_1 *is obtained and considering this element as the outcome of the composite experiment.*

According to property d) of section 2 E' and E" are equivalent and thus we ought to use the same model for both of them. But according to rule 1 the model for E' is a symmetric probability space with Ω_1 as space of events and with the Laplace-definition. This means that we should also use this model for E" and this is exactly what happens. The notation "$|\Omega_1$" is used to indicate the conditioning on Ω_1 in either of the two ways indicated by E' or E". The Laplace-definition applied to E' now leads straight to (2), for according to this definition we have

$$P(\Omega_2|\Omega_1) = N(\Omega_1 \cap \Omega_2)/N(\Omega_1) =$$
$$= \{N(\Omega_1 \cap \Omega_2)/N\}/\{N(\Omega_1)/N\} =$$
$$= P(\Omega_1 \cap \Omega_2)/P(\Omega_1),$$

where the unconditional probabilities pertain to one random drawing from Ω. Note that neither E' nor E" can be performed if $N(\Omega_1) = 0$; thus the reasoning only holds if $P(\Omega_1) > 0$.

The equivalence of E' and E" seems rather evident but the following anecdote shows that this does not hold for everybody*). An advertising agency organized a quiz in order to promote some product. The quiz consisted of some simple questions and the response was overwhelming. Thousands of answers were received and, of course, the prizes had to be awarded at random among the correct solutions. To this end the agency hired a number of working students in order to sift out the wrong answers (which were comparatively few). This took several weeks time and when this work was completed the winners were drawn at random from the correct solutions. This procedure corresponds to E' and it is perfectly correct. How much more simple and less time-consuming it would have been, however, to use procedure E"!

The generalization of (2) to partitioned probability spaces is straightforward. We will skip it. It is also clear that from (2) the *general multiplication law* and the theorem on composite probabilities follow and that statistical independence means that conditional probabilities are equal to the corresponding unconditional ones.

After these preparations we want to formally introduce the use of conditional probabilities in building up models for stepwise experiments. Let $E^{(1)}$ be an experiment with $\Omega^{(1)}$ as its space of events and $P^{(1)}$ as its probability function on $\Omega^{(1)}$, all according to previous rules. Let $E^{(2)}$ be a second experiment with space of events $\Omega^{(2)}$, but depending on the result of $E^{(1)}$ in the following sense: for every $\omega^{(1)} \in \Omega^{(1)}$ an experiment $E^{(2)}_{\omega^{(1)}}$ is given which has a probability function $P^{(2)}_{\omega^{(1)}}$ (on $\Omega^{(2)}$), depending on $\omega^{(1)}$, again in accordance with previous rules. The composite experiment $E = (E^{(1)}, E^{(2)})$ is composed of $E^{(1)}$ and $E^{(2)}_{\omega^{(1)}}$, where $\omega^{(1)}$ is the event realized in $E^{(1)}$. In these circumtances $E^{(2)}$ is called *statistically dependent* on $E^{(1)}$ and E is called a *stepwise composed experiment*. By induction

---

*) H. Piller, personal communication.

we get any finite number of steps.

The model for E must be in accordance with previous model-rules and to find such a model we again consider two equivalent experiments:

E': *one realisation of* $E^{(2)}_{\omega^{(1)}}$ *for given* $\omega^{(1)}$*; the probability for obtaining* $\omega^{(2)} \in \Omega^{(2)}$ *is then*

$$P^{(2)}_{\omega^{(1)}}(\omega^{(2)}).$$

E": *repeating* E *until* $E^{(1)}$ *gives* $\omega^{(1)}$ *and looking at the result of* $E^{(2)}$ *in that trial. This means:imposing the condition* $\omega^{(1)}$ *on* $\vec{E}$*, and thus the model for* E *must be such that the probability for obtaining* $\omega^{(2)} \in \Omega^{(2)}$ *is*

$$P(\omega^{(2)}|\omega^{(1)}).$$

The equivalence of E' and E" now leads to

$$(3) \qquad P(\omega^{(2)}|\omega^{(1)}) = P^{(2)}_{\omega^{(1)}}(\omega^{(2)})$$

and together with the multiplication law, which must also hold in the model for E, we find that we have to build up this model on $\Omega^{(1)} \times \Omega^{(2)}$ by means of

$$(4) \qquad P((\omega^{(1)},\omega^{(2)})) = P^{(1)}(\omega^{(1)})P^{(2)}_{\omega^{(1)}}(\omega^{(2)}).$$

This is the only possibility if we want to obey our previous rules and the principle of equivalence. This result can be summarized as follows.

Rule 5. *For stepwise experiments where for every step previous rules lead to a probability space depending on the results of previous steps, a model is built up by means of conditional probability spaces for the steps and by means of the multiplication law for simultaneous probabilities.*

8. CONDITIONAL PROBABILITIES AND INFORMATION

Although some details were glossed over in section 7 the treatment of a seemingly obvious method may seem rather extensive to some readers. But one must be careful as the following example is meant to show. A player throws a good six-sided die and you are to guess the result. This is the situation of section 2: your guess does not really matter as long as it is

Lack of knowledge about the information-policy can be incorporated adequately in the model by introducing an unknown partitioning of Ω, i.e. unknown conditions for the conditional probabilities. In our example with the die this would, in case "it is not 6", lead to five possible outcomes with unknown probabilities, which can have only a finite number of different values because there are only a finite number of possible partitionings. Further knowledge about the actual values but also about the actual information-policy could then be gathered by observing repeated independent trials of the same experiment. On the other hand it may be remarked that one may remedy the situation by randomizing ones guess among the numbers $1,\ldots,5$. Then at least the probability of a right guess is 1/5. Thus perhaps, one should never read a newspaper without a die or a coin at hand.

9. FINAL REMARKS

Although up till this point we only have finitely many rational probabilities in a probability space the generalization to infinitely many real ones and to continuous probability spaces is of a less fundamental nature. It is all passing to the limit and approximating discrete situations by means of continuous ones for the sake of mathematical convenience and greater generality. So we need not be sorry that the scope of this paper does not allow us to go over all that. It is a pity that the space allotted is too small to talk about some other things like: the interpretation of probabilities in order to go back from the model to reality after the analysis in the model is completed and to the phenomenon that statisticians do not only seek to predict the future, but also the past: the example in section 7 is of that character just as e.g. the method of confidence intervals for unknown parameters. These things are interesting but they will have to wait.

REFERENCES

D. VAN DANTZIG (1957), *Statistical priesthood (Savage on personal probabilities)*, Stat. Neerl. 11, 1-16.

J. HEMELRIJK (1968), *Back to the Laplace definition*, Stat. Neerl. 22, 13-21.

B. VAN ROOTSELAAR and J. HEMELRIJK (1969), *Back to "Back to the Laplace definition"*, Stat. Neerl. 23, 87-89.