X.

THE FOUNDATION OF THE THEORY OF PROBABILITY—II.[1]

[From the Dublin Institute for Advanced Studies.]

By ERWIN SCHRÖDINGER.

[Read 9 DECEMBER, 1946. Published 30 JUNE, 1947.]

To introduce the product-rule as an axiom, as we did in sect. 5, is unsatisfactory. It is anything but self-evident. Indeed it could not be that. By adopting it straight away, we surreptitiously adopt a certain normalisation of the "measure of likelihood," a change of which would patently upset the rule. We shall investigate here how far we can get in specifying this normalization openly. It means completing the one contained in the "second convention," sect. 4, where we had emphasized at the end that it was only a first step, capable of being consummated. By "how far we can get" I mean to say: in the way of splitting, as it were, the axiom in one part that is mere convention, and another one that, if it be deemed axiomatic, at least stands to reason.

Since the following considerations branch off from the end of sect. 4, they must ignore the later parts of the first paper: neither the product rule nor the summation rule must be made use of.

8. Establishing a Lemma.

If an event β can be exhaustively analysed into s neatly separated (non-overlapping, mutually exclusive) different "manners" in which it can happen, which as "events" may be called $\beta_1, \beta_2, \ldots \beta_s$, then the conjectures we may make, given a certain state of our knowledge, about the coming true of the events $\beta_1, \beta_2, \ldots \beta_s$ severally, constitute, when taken together, a conjecture about the event β, under the same state of knowledge. Moreover the latter conjecture will be at least as definite as the

[1] See Proc. Roy. I. Acad. 51 (A), 51, 1947.

vaguest of the s constituting conjectures. In particular, if our knowledge allows us to associate numerical probabilities $p_{\beta_1}, p_{\beta_2}, \ldots p_{\beta_s}$ with $\beta_1, \beta_2, \beta_3, \ldots \beta_s$ separately, this must result in associating also a definite p_β with the event β.

The numerical value of p_β can in this case only depend on the numerical values of the p_{β_i}, not on the particular nature of the β_i. Hence it must depend on them in a universal and symmetrical way. Moreover (from the first convention, p. 55) p_β must increase if any particular p_{β_i} is increased while the others are kept constant. Thus we must have

$$p_\beta = L_s(p_{\beta_1}, p_{\beta_2}, \ldots p_{\beta_s})$$

where every $L_s (s = 2, 3, 4 \ldots)$ is a universal symmetric function that increases monotonically with each of its s arguments.

We call this our *Lemma*. Somebody who refuses to admit the stringency of our above *deduction* may prefer to call it an axiom. At any rate I do not think much would be gained by trying to analyse it further into simpler statements.

As an obvious Corrollary we add, that

$$L_s(x, 0, 0, \ldots 0) = x \tag{14}$$

for any s.

9. Completing the Normalization.

Now envisage n mutually exclusive events $\alpha_1, \alpha_2, \ldots \alpha_n$ of which the logical[2] sum $\alpha_1 + \alpha_2 + \ldots + \alpha_n$ constitutes the nil-event, thus

$$p_{\alpha_1 + \alpha_2 + \ldots \alpha_n} = 1\,.$$

From our Lemma we draw

$$\begin{aligned} p_{\alpha_1 + \alpha_2} &= L_2(p_{\alpha_1}, p_{\alpha_2}) \\ p_{\alpha_1 + \alpha_2 + \alpha_3} &= L_3(p_{\alpha_1}, p_{\alpha_2}, p_{\alpha_3}) \\ &\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot \\ p_{\alpha_1 + \alpha_2 + \ldots + \alpha_{n-1}} &= L_{n-1}(p_{\alpha_1}, p_{\alpha_2}, \ldots p_{\alpha_{n-1}}) \\ 1 &= L_n(p_{\alpha_1}, p_{\alpha_2} \ldots p_{\alpha_n})\,. \end{aligned} \tag{15}$$

[2] If you object to the term "logical" call it "factual." It means: either α_1 or $\alpha_2 \ldots$ or α_n comes true (at least; but our events are exclusive).

We first fix our attention to the last equation. It is easily seen that it has one and only one solution for which

$$p_{\alpha_1} = p_{\alpha_2} = \ldots = p_{\alpha_n} .$$

Indeed (from the Corrollary (14))

$$L_n(g, g, \ldots g) \tag{16}$$

is zero for $g = 0$. It increases monotonically with g and must reach the value 1 before g does. For if we had

$$L_n(1, 1, 1 \ldots 1) < 1$$

then a fortiori

$$L_n(1, 0, 0 \ldots 0) < 1$$

But the latter *equals* 1, from the Corrollary.

We call the value of g for which (16) reaches unity g_1, moreover we call

$$L_s(g_1, g_1, \ldots g_1) = g_s, \quad (s = 2, 3, \ldots n - 1) .$$

Then we have, in that case,

$$\begin{aligned} p_{\alpha_1} &= g_1 \\ p_{\alpha_1 + \alpha_2} &= g_2 \\ p_{\alpha_1 + \alpha_2 + \alpha_3} &= g_3 \\ \cdot \quad \cdot \quad \cdot \quad & \cdot \quad \cdot \quad \cdot \quad \cdot \\ p_{\alpha_1 + \alpha_2 + \ldots + \alpha_{n-1}} &= g_{n-1} \end{aligned}$$

In this sequence every subsequent line refers to an event that is included in the previous one but is obviously not equivalent to it. Hence (from the first convention, p. 55) we have:

$$0 < g_1 < g_2 < \ldots < g_{n-1} < 1 .$$

Hence by a universal monotonical transformation of our "measure of likelihood" p,

$$p' = f(p)$$

we can shift $g_1, g_2, \ldots g_{n-1}$ to $\frac{1}{n}, \frac{2}{n}, \ldots \frac{n-1}{n}$ respectively. We need only demand

$$f(g) = \frac{s}{n} \qquad (s = 1, 2, 3 \ldots n - 1) .$$

A careful consideration shows that this normalizing procedure can be applied successively to $n = 2^N$ with $N = 1, 2, 3, 4 \ldots$ in inf., no subsequent steps interfering with the preceding ones. In this way the normalization seizes ultimately on every rational point.

The normalization of sect. 4 amounted to the first step $(N = 1)$, supplemented by a certain symmetrization of the domains $p < \frac{1}{2}$ and $p > \frac{1}{2}$, by which the theorem (2) (about complementary probabilities) was obtained at one go. It is easy to see that this symmetrization too is not destroyed by the later steps.

We formulate thus our

3RD CONVENTION: *The probability p_s of the logical sum of any s out of n mutually exclusive, equally probable events with logical sum nil, can and shall, by normalization be allotted the numerical value*

$$p_s = \frac{s}{n}.$$

10. Proof of the product rule in a special case.

Let

$$a_1, \ a_2, \ \ldots \ a_n \tag{16}$$

be n mutually exclusive, *equally probable* events of logical sum nil. Construct from them the events

$$\alpha = a_1' + a_2' + \ldots a_k'$$
$$\beta = a_1'' + a_2'' + \ldots a_l''$$

where the a_i' are any k events, different from each other, out of the set (16), and the a_i'' are any l events, different from each other, out of the same set (16). Then the logical product $\alpha\beta$ is

$$\alpha\beta = \hat{a}_1 + \hat{a}_2 + \ldots + \hat{a}_q,$$

where the $\hat{a}_i$ are the q events, different from each other, that belong both to the a_i' and to the a_i''. From the third convention

$$p_\alpha = \frac{k}{n} \qquad p_\beta = \frac{l}{n} \qquad p_{\alpha\beta} = \frac{q}{n}.$$

Hence

$$\begin{aligned} p_{\alpha\beta} &= \frac{k}{n}\frac{q}{k} = p_\alpha \frac{q}{k} \\ &= \frac{l}{n}\frac{q}{l} = p_\beta \frac{q}{l}. \end{aligned} \tag{17}$$

Now what is $\frac{q}{k}$? If we received the additional information that α comes true, this information (being symmetrical with respect to the a_i') leaves their probabilities *equal*, but changes them from $\frac{1}{n}$ to $\frac{1}{k}$, since α has now become the nil-event. Moreover with the additional knowledge β becomes

$$\beta = \hat{a}_1 + \hat{a}_2 + \ldots + \hat{a}_q.$$

Hence, in the notation explained in sect. 5,

$$p_\beta(\alpha^+) = \frac{q}{k}.$$

In the same way

$$p_\alpha(\beta^+) = \frac{q}{l}.$$

So (17) reads

$$\begin{aligned} p_{\alpha\beta} &= p_\alpha \, p_\beta(\alpha^+) \\ &= p_\beta \, p_\alpha(\beta^+). \end{aligned} \tag{18}$$

This proves the product rule in our special case.

Write the first eqn. (18) once again as an identify

$$\frac{q}{n} = \frac{k}{n}\frac{q}{k}.$$

Within limits that stand to reason according to the first convention and the meaning these fractions have as measures of likelihood any two of them can be chosen arbitrarily by choosing the numbers n, k, q appropriately. In other words, one can always construct an "urn-example" in which two of the three probabilities appearing in the first equation (18) have any values prescribed in accordance with the first convention *cum* reason.[3] The third probability is then—for the urn-model—determined by that equation.

11. The General Product-Rule.

Now envisage two arbitrary events α, β and suppose that under a given knowledge, two of the following three probabilities can be determined:

$$p_{\alpha\beta}, \qquad p_\alpha, \qquad p_\beta(\alpha^+) \tag{19}$$

Then according to what has been said at the end of the previous section, we can construct a model couple of urn-events α', β' such that of the following three equations

$$p_{\alpha\beta} = p_{\alpha'\beta'}, \qquad p_\alpha = p_{\alpha'}, \qquad p_\beta(\alpha^+) = p_{\beta'}(\alpha'^+)$$

two hold, while the remaining one—nay even the existence of its first member—is to be established.

Now the queried equation could fail to hold in either of two ways, viz.

(i) its first member exists, but instead of equality there is inequality ($\gtrless$).

(ii) its first member does not exist.

[3] I mean to say, e.g. the demand $p_{\alpha\beta} = \frac{1}{5}$, $p_\beta(\alpha^+) = \frac{1}{10}$ is unreasonable, with the first convention.

The first submission can, I think, be rejected on the strength of the First Convention (sect. 4, p. 55), according to which an event that is less likely than another one has a smaller numerical probability and vice versa. But we must now separate the cases.

If the second equation is in question, while the first and third one hold, we say: if p_α were $\gtrless p_{\alpha'}$, this would mean that α is $\genfrac{}{}{0pt}{}{\text{more}}{\text{less}}$ likely than α'; this together with the inference drawn from the last equation, that the likelihood of β, in case α happens, is the same as that of β' happening, in case α' happens—must $\genfrac{}{}{0pt}{}{\text{increase}}{\text{reduce}}$ the likelihood of both α and β happening as against the likelihood of both α' and β' happening; so we would have to have $p_{\alpha\beta} \gtrless p_{\alpha'\beta'}$, in contradiction to the assumption, that they are equal.

If the third equation is in question, the argument is similar, we need not repeat it.

The third case (first equation doubtful, the other two holding) ought, I think, to be reduced to one of the previous cases by potentially modifying the urn-example, thus: if $p_{\alpha\beta}$ were $\gtrless p_{\alpha'\beta'}$, it would certainly be possible to make them equal by keeping $p_{\beta'}(\alpha'^{+})$ constant and $\genfrac{}{}{0pt}{}{\text{increasing}}{\text{reducing}}$ $p_{\alpha'}$ $\genfrac{}{}{0pt}{}{\text{beyond}}{\text{below}}$ p_α; indeed $p_{\alpha\beta}$ must lie between 0 and $p_\beta(\alpha^{+})$ thus between 0 and $p_{\beta'}(\alpha'^{+})$, and $p_{\alpha'\beta'}$ can reach both these limits, viz. for $p_{\alpha'} = 0$ and $p_{\alpha'} = 1$ respectively; in this manner the third case is reduced to the first one and it follows, that actually the potential modification of the urn-example amounted to nothing.

If one grants this ratiocination on the strength of the first convention, then, to establish the general product rule, nothing more is needed, than the *axiomatic postulate*:

If our knowledge suffices definitely to exclude for the numerical probability of a certain event all values save one, but definitely does not exclude that one, we admit it to have this value.

If one does not agree with our above deduction "from pure reason," one has to go a little further and admit axiomatically, that between the three probabilities (19) there is a universal dependence which does not depend on the nature of the two events in question.

Even this is much less than to admit the quantitative product-rule straight away, as we did in sect. 5.