VOL. LIII. NO. 210.] [April, 1944.

MIND

A QUARTERLY REVIEW

OF

PSYCHOLOGY AND PHILOSOPHY

I.—HR. VON WRIGHT ON THE LOGIC OF INDUCTION (II.).

BY C. D. BROAD.

(III) ATTEMPTS TO JUSTIFY INDUCTION *A POSTERIORI*.

(B) PROBLEMATIC.

AT the end of the first part of this paper we had seen that, for reasons given by Hr. von Wright which seem conclusive, it is idle to hope that an inductive generalisation which is genuinely synthetic can ever be established with certainty by deductive reasoning from instantial premisses alone or combined with postulates about nature. We have now to consider whether it is possible to show by means of the principles of Probability that, under certain conditions, instantial premisses can render an inductive generalisation highly probable. Even if this can be shown we shall not be at the end of our troubles. For the meaning and implications of statements of the form 'p has such and such a degree of probability given h' are not immediately obvious or universally agreed upon. It would remain therefore to decide what interpretation to put on such statements; and to consider what relevance, if any, the fact that an inductive generalisation was highly probable with respect to certain instantial data would have to expectation and to action. Therefore, as Hr. von Wright points out, there are two problems, one of formal analysis and one of interpretation. We will now consider them in turn.

(1) *Formal Analysis of Inductive Probability.*—Let p and h be any two propositions. Consider the as yet undefined expression 'p has with respect to h a probability of degree x'. The form of this expression tells us that 'probability' is understood to be a relation which one proposition may bear to another, and that this relation has magnitude; but it tells us nothing further. We therefore proceed to lay down a set of postulates which together govern the use and the transformations of such expressions but do not commit us to any particular analysis of their meaning.

Postulates.—Hr. von Wright gives six postulates for this purpose. I shall take as the first the one which he takes as fourth. They run as follows.

(i) A given p has with regard to a given h only one degree of probability. (On the basis of this we can talk of '*the* probability of p with respect to h'. I shall symbolise this, as Johnson and Keynes do, by the symbol p/h. So the statement that 'the probability of p with respect to h is x' will be written in the form of the equation $p/h = x$, which might be compared with such an equation as $dy/dx = z$ in the differential calculus. Following Johnson I shall call the proposition on the left of the solidus the 'proposal' and that on the right of it the 'supposal'.)

(ii) The possible numerical values of such expressions as p/h are all the real numbers from 0 to 1, both inclusive.

(iii) If h implies p, then $p/h = 1$.

(iv) If h implies not-p, then $p/h = 0$. (It must not be assumed that the converse of this or of the preceding postulate is true.)

(v) $p \,\&\, q/h = p/h \times q/p \,\&\, h = q/h \times p/q \,\&\, h$.

(vi) $p \vee q/h = p/h + q/h - p \,\&\, q/h$.

The last two postulates, which may be called respectively the *Conjunctive* and the *Disjunctive* postulate, enable us to express the probability of a conjunctive or a disjunctive proposal, with respect to a given supposal, in simpler terms. It should be noted that, unless certain special conditions are fulfilled, neither $p \,\&\, q/h$ nor $p \vee q/h$ can be expressed wholly in terms of p/h and q/h. The necessary and sufficient condition for this is, in the case of the Conjunctive Postulate, that $q/p \,\&\, h = q/h$ and (what is entailed by this) that $p/q \,\&\, h = p/h$. In the case of the Disjunctive Postulate the condition is that $p \,\&\, q/h = 0$. The first condition is that p and q are *independent* with respect to h; the second is that they are *exclusive* with respect to h.

Lemmas.—In order to make use of these Postulates we shall constantly need certain immediate consequences of them. I shall therefore proceed to prove these in a series of lemmas.

Lemma I. $\overline{p}/h = 1 - p/h$.
For, since the disjunctive proposition $p \vee \overline{p}$ is known to be true, it is implied by any proposition h. Therefore, by Postulate (iii), $p \vee \overline{p}/h = 1$. But p and $\overline{p}$ are mutually exclusive. Therefore, by Postulate (vi), $p \vee \overline{p}/h = p/h + \overline{p}/h$, which proves the proposition.

Lemma II. If $p \supset q$, then $q/p \,\&\, r = 1$, whatever r may be. For, if $p \supset q$, then $p \,\&\, r \supset q$, whatever r may be. Therefore, by Postulate (iii), $q/p \,\&\, r = 1$.

Lemma III. If two propositions, p and q, are logically equivalent, *i.e.*, if each implies the other, then their probabilities with respect to any proposal h are equal.

From Postulate (v) and Lemma II it follows that, if p implies q, then $p \,\&\, q/h = p/h$. From the same premisses it follows that, if q implies p, then $p \,\&\, q/h = q/h$. Therefore, if p is equivalent to q, $p/h = q/h$.

Lemma IV. If $q_1, q_2, \ldots q_n$ are a set of n mutually exclusive alternatives, and if p implies the disjunction of them, then $p/h = \overset{r=n}{\underset{r=1}{\mathrm{S}}} q_r/h \times p/q_r \,\&\, h$.

In general if $p \supset q$ then $p \equiv p \,\&\, q$. Therefore in the present case

$$p \equiv : p \,\&\, q_1 \,.\vee.\, p \,\&\, q_2 \,.\vee \ldots p \,\&\, q_r \,.\vee \ldots p \,\&\, q_n.$$

Now the alternatives on the right of this equivalence are mutually exclusive. Therefore, by Postulate (vi) and Lemma III,

$$p/h = \overset{r=n}{\underset{r=1}{\mathrm{S}}} p \,\&\, q_r/h.$$

And, by Postulate (v), $p \,\&\, q_r/h = q_r/h \times p/q_r \,\&\, h$; which proves the proposition.

Lemma V. If $q_1, q_2, \ldots q_n$ are not only mutually exclusive but also collectively exhaustive, then, whatever p may be, $p/h = \overset{r=n}{\underset{r=1}{\mathrm{S}}} q_r/h \times p/q_r \,\&\, h$.

For the disjunction of all the q's is now a true proposition, and so it is implied by every proposition and therefore by p whatever p may be. Hence the conclusion follows as in Lemma IV.

Lemma VI. $$p/q \,\&\, h = \frac{p/h \times q/p \,\&\, h}{q/h}.$$

This is an immediate consequence of Postulate (v). It may be called the *Principle of Inverse Probability.*

Lemma VII. If $q_1, q_2, \ldots q_n$ are a set of mutually exclusive alternatives, and if p implies the disjunction of them, then

$$q_r/p \,\&\, h = \frac{q_r/h \times p/q_r \,\&\, h}{\overset{r=n}{\underset{r=1}{\mathrm{S}}}\, q_r/h \times p/q_r \,\&\, h}.$$

This is an immediate consequence of Lemmas IV and VI. It may be called the *Bayes Principle.* If each of the alternatives implies p, all terms of the form $p/q_r \,\&\, h$ become equal to 1, and the Principle takes the simplified special form

$$q_r/p \,\&\, h = \frac{q_r/h}{\overset{r=n}{\underset{r=1}{\mathrm{S}}}\, q_r/h}.$$

Lemma VIII. If $q_1, q_2, \ldots q_n$ are a set of mutually exclusive alternatives, and if p implies the disjunction of them, and if $q_{r_1} \ldots q_{r_k}$ are a selection of k of them, then

$$q_{r_1} \vee q_{r_2} \vee \ldots q_{r_k}/p \,\&\, h = \frac{\overset{s=k}{\underset{s=1}{\mathrm{S}}}\, q_{r_s}/h \times p/q_{r_s} \,\&\, h}{\overset{r=n}{\underset{r=1}{\mathrm{S}}}\, q_r/h \times p/q_r \,\&\, h}.$$

This is an immediate consequence of Lemma VII and Postulate (vi). It may be called the *Extended Bayes Principle.* If each of the alternatives implies p it simplifies in the same way and for the same reason as the Bayes Principle.

Theorems connecting Probability with Induction.—Hr. von Wright proceeds to prove from his postulates a number of theorems about the probability of inductive generalisations with respect to instantial propositions, and to state the conditions under which they hold.

My own experience is that I can see best what such theorems and their conditions really amount to when I exemplify them by concrete illustrations of drawing counters from bags, throwing dice, etc. I suspect that many others will be in the same position. I propose therefore to begin in each case with a concrete example and then to generalise from it. In each theorem I shall conduct the proof in my own way, which may or may not be exactly that followed by Hr. von Wright; and I carry the argument up to the point where nothing further is needed but an application of pure mathematics which the reader will be asked to take on trust.

Before going further it is desirable to make the following remarks about notation. In some of our theorems we have to consider the probability of a proposition with respect to a supposal which includes data about the probability of another proposition. *E.g.*, we may ask: What is the probability of drawing three white counters in succession from a bag on the supposition that the probability of drawing a white counter on each occasion is so-and-so ?

We need some system of bracketing which will, *e.g.*, clearly distinguish the following two entirely different propositions. (i) ' The probability of p, on the supposition that the probability of q-and-r with respect to h is x, is y.' And (ii) ' The probability of p, on the supposition that q is true and that the probability of r with respect to h is x, is y '. Now this might be done by using a combination of round, square, and curly brackets ; but this would be extremely clumsy. Instead I shall make use of dots according to the following convention. When the main supposal contains data about probability the first solidus will be immediately followed by one or more dots. These dots will constitute the beginning of a bracket, and this will be closed by the first occurrence of the *same number* of dots further to the right of the expression. Thus, *e.g.*, the first of the two propositions enunciated above would be expressed by the formula

$$p/\,.\,q\ \&\ r/h = x\,.=\,y.$$

And the second by the formula

$$p/:q\,.\,\&\,.\,r/h = x:=\,y.$$

We are now in a position to deal with the Theorems.

Theorem 1. *Bernoulli's Theorem.* (1·1) *Direct Principle of Greatest Probability.*—Suppose that there is a bag containing n counters, of which exactly m are white. A sequence of N trials is made according to the following rule. On each occasion one counter is drawn, the colour is noted, the counter is then replaced, and the contents of the bag are well stirred up before the next trial is made. These rules ensure the fulfilment of the following conditions :—

(i) The sequence of trials is in principle indefinitely extensible.

(ii) The probability that the counter drawn at the $r + 1$-th trial will be white is independent of the frequency and the order with which white drawings have occurred among the previous r trials.

(iii) The probability that the counter drawn will be white on any one occasion is the same as the probability that it will be white on any other occasion. (In our example it is m/n on every occasion.)

The second and third of these conditions may be called the *Bernoullian Conditions*. The bearing of the three conditions is made plain by considering examples in which they are not fulfilled. If, *e.g.*, the rule were that the counters drawn are not to be replaced, all three conditions would break down. The sequence of drawings would end at the n-th term. The probability that the $r+1$-th trial will give a white counter will depend on the number of white counters which have been drawn in the previous r trials. If, *e.g.*, p have been white, the bag will contain $n-r$ counters, of which $m-p$ are white, at the time when the $r+1$-th trial is to be made; so the probability that this trial will give a white result is $\frac{m-p}{n-r}$. Suppose, again, that the drawings were to be made alternately from two bags, each containing n counters, one of which contained m_1 whites and the other m_2 whites. Suppose that the counters were to be replaced after each drawing. Then conditions (i) and (ii) would hold, but condition (iii) would break down.

We ask ourselves the question: Given that the Bernoullian conditions hold, what is the most probable proportion of whites in a sequence of N trials? The answer is that the most probable *number* of whites is the nearest integer to $\frac{m}{n} \times N$, and therefore the most probable *proportion* of whites is this integer divided by N.

It remains to generalise this proposition and to prove it.

Let Q be any characteristic present in every term of a series of trials which can be extended indefinitely. Let R be another characteristic (or a determinate form of Q) which may be present in any proportion of the instances of Q. Let us assume (*a*) that the probability of an instance of Q being R is independent of the number and distribution of instances of R among the previous instances of Q; and (*b*) that the antecedent probability of any instance of Q being R is the same, *viz.*, p. We will symbolise these two suppositions by the conjunction $h\,.\,\&\,.\,\mathrm{R}(x)/\mathrm{Q}(x)\,\&\,h=p$. Let us denote the relative frequency of R's among the first N instances of Q by $f_N(\mathrm{R}\,;\,\mathrm{Q})$. Then

$$f_N(\mathrm{R}\,;\,\mathrm{Q}) = \frac{r}{N}\,.\,/:h\,.\,\&\,.\,\mathrm{R}(x)/\mathrm{Q}(x)\,\&\,h=p$$

is a maximum when r is the nearest integer to pN. This is the *Direct Principle of Greatest Probability*.

The proof is as follows. The proposition $f_N(\mathrm{R}\,;\,\mathrm{Q}) = \frac{r}{N}$ is the same as the proposition that there have been r instances of

R among the first N instances of Q. Now these r instances of R might have been distributed among the N instances of Q in ${}^{N}C_{r}$ different ways. So the proposition in question is equivalent to the disjunction of ${}^{N}C_{r}$ mutually exclusive alternatives, each of which is a conjunction of N conjuncts. A typical one of these alternatives would be the proposition

$$\mathrm{R}(x_1) \,\&\, \ldots\, \mathrm{R}(x_r) \,\&\, \bar{\mathrm{R}}(x_{r+1}) \,\&\, \ldots\, \bar{\mathrm{R}}(x_{\mathrm{N}}),$$

which represents the possibility that all the r R's come first and then are followed by all the $N - r$ non-R's. Since the alternatives are mutually exclusive, the probability that $f_{\mathrm{N}}(\mathrm{R}\,;\,\mathrm{Q}) = \frac{r}{N}$ with respect to the assumed data is equal, by Postulate (vi), to the sum of the probabilities of the several alternatives with respect to the same data. Now these data include the Bernoulli conditions, in accordance with which the probability of any instance of Q being R is the same as that of any other instance being R and is independent of the number and the distribution of R's among the preceding Q's. Therefore, by Postulate (v) and Lemma I, the probability with respect to the assumed data of each of the conjunctive alternants is the same, *viz.*, $p^r(1-p)^{\mathrm{N}-r}$. Therefore the sum of these probabilities is ${}^{\mathrm{N}}C_r p^r(1-p)^{\mathrm{N}-r}$. So we have proved that

$$f_{\mathrm{N}}(\mathrm{R}\,;\,\mathrm{Q}) = \frac{r}{N} \,.\, /\, :\, h \,.\, \&\,.\, \mathrm{R}x/\mathrm{Q}x \,\&\, h = p \,:=\, {}^{\mathrm{N}}C_r p^r(1-p)^{\mathrm{N}-r}.$$

It is then a matter of elementary algebra to show that ${}^{\mathrm{N}}C_r p^r(1-p)^{\mathrm{N}-r}$ is a maximum for a fixed value of N when r is the nearest integer to pN; and this is what we set out to prove.

(1·2) *Direct Principle of Great Numbers.*—This is the second part of the Bernoulli Theorem. It may be stated accurately in words as follows. Let δ and ϵ be any two pre-assigned quantities. Then, no matter how small they may be, there is a series of Q's such that for every longer series of Q's than this the probability that the proportion of R's in it will not differ from p by more than δ does not differ from 1 by more than ϵ.

The proposition can be stated more colloquially but less accurately as follows. By continuing the series of Q's far enough you can always ensure that the probability of the proportion of R's in it differing by as little as you please from p will differ by as little as you please from 1. The accurate expression for this proposition in our symbolism is

$$(\delta, \epsilon) \,:\,:\, (\exists n) \,:\,: N > n \;\supset_{\mathrm{N}}\, : f_{\mathrm{N}}(\mathrm{R}\,;\,\mathrm{Q}) \doteq p \pm \delta \,.\, /\, :\, h \,.\, \&\,.\, \mathrm{R}x/ \mathrm{Q}x \,\&\, h = p \,:\, \geqslant 1 - \epsilon.$$

The proof of this depends on the following two propositions of pure mathematics. (i) *Stirling's Theorem* that when n is large $n!$ approximates to $n^n e^{-n}\sqrt{2\pi n}$. (ii) The fact that the sum of a very large number of very small terms can be expressed as the integral of a function between two limits. The argument is as follows.

To say that the relative frequency of R's in a series of N Q's does not fall outside a certain range $\frac{M \pm r}{N}$ is equivalent to saying that it has one or other of the $2r + 1$ possible values

$$\frac{M-r}{N}, \frac{M-r+1}{N}, \ldots \frac{M}{N}, \frac{M+1}{N}, \ldots \frac{M+r}{N}.$$

So the proposal $f_N(\mathrm{R}\,;\mathrm{Q}) \doteq p \pm \delta$ is in fact a disjunctive proposition in which the various alternants are mutually exclusive. Therefore, by Postulate (vi), its probability with respect to the supposal is equal to the sum of the probabilities of the several alternants with respect to the same supposal. Now these are already known from the first part of the Theorem. If N is very great, the number of alternatives included in even a small interval δ will be very great. For the number is $2r + 1$ where $r = N\delta$. Again, the probability of the proportion of R's being exactly any one of these alternative fractions is very small. Therefore we have the necessary and sufficient conditions for replacing the sum by an integral. Again, since the probabilities of the several alternative proportions of R's among the N Q's, as determined by the first part of the Theorem, are all of the form

$$\frac{N!}{(M+k)!(N-M-k)!} p^{M+k}(1-p)^{N-M-k}$$

where both N and M are large, we can apply Stirling's Theorem to them.

As a result of these two considerations we can show by mere mathematical manipulation that as N increases

$$f_N(\mathrm{R}\,;\mathrm{Q}) \doteq p \pm \delta \,./:\, h \,.\,\&\,.\, \mathrm{R}x/\mathrm{Q}x \;\&\; h = p$$

approaches $$\frac{1}{\sqrt{2\pi N p q}} \int_{-N\delta}^{N\delta} e^{-\frac{x^2}{2Npq}} dx$$

where we write q as an abbreviation for $1 - p$. Now it is easy to show that the latter expression is equal to $\frac{2}{\sqrt{\pi}} \int_0^{\sqrt{\frac{N\delta}{2pq}}} e^{-x^2} dx$.

Now, however small δ may be, the upper limit of this integral can be made as large as we please by sufficiently increasing N. And it is known that the integral can be made to differ from $\frac{1}{2}\sqrt{\pi}$ by as little as we please by making its upper limit large enough. Therefore, however small δ and ϵ may be, the expression as a whole can be made to differ from 1 by less than ϵ if N be sufficiently increased. And this is what we had to prove.

Theorem 2. *The Inverse Bernoulli Theorem.* (2·1) *Inverse Principle of Greatest Probability.*—The theorem with which we are now to be concerned is called by Hr. von Wright *Bayes's Theorem.* I prefer the name which I have given to it. Certainly this theorem depends upon our Lemma VII, which I have called *Bayes's Principle;* but the same may be said of almost all applications of inverse probability.

In the Bernoulli Theorem the proposal was a statistical proposition, and the supposal contained a probability proposition. In the Inverse Bernoulli Theorem the proposal is a probability proposition, and the supposal contains a statistical proposition. So we are now concerned with the probability of a probability having a certain value, given that a frequency has a certain value. This is not an easy notion, and it is particularly important to make its meaning plain by means of an example of drawing counters from bags.

Suppose that there are a number of bags, each containing n counters. We will call them B_1, B_2, . . . B_r, . . . B_z. We are told that the first contains m_1 white counters, the second m_2 white counters, and so on. We are told that someone has made N drawings from *one and only one* of these bags and that the Bernoulli conditions have been fulfilled, *i.e.*, he has replaced the counter after each drawing in the bag from which he drew it and has stirred the counters well before drawing again. What we are *not* told is the *particular bag* from which the drawings have been made.

If the bag used happened to be B_1 the probability at each drawing that the counter drawn would be white was $\frac{m_1}{n}$; if it happened to be B_2 this probability was $\frac{m_2}{n}$; and so on. So what might be called 'the antecedent probability that the antecedent probability of drawing a white was $\frac{m_r}{n}$' is simply the antecedent probability that the bag from which the drawings were made was B_r, *i.e.*, the one in which the proportion of white counters is $\frac{m_r}{n}$.

Now this probability *might* be given to us by the rules of the game. Suppose, *e.g.*, that we were told that there was, in addition to the z bags already mentioned, another bag containing n_1 counters marked '1', n_2 counters marked '2', and so on, thoroughly mixed up. We are told that the experimenter first drew a single counter from this bag and then made his drawings from the bag which bears the number marked on the counter which he had drawn. The probability, with respect to this rule, that he would have drawn a counter marked with the number 'r' is of course

$$\frac{n_r}{n_1 + n_2 + \ldots n_z}.$$

Now, given the rule, this is the probability that he will make his experiments on bag B_r. And, given the information about the constitution of the contents of the bags, this is the probability that the probability of drawing a white counter on each occasion throughout his experiment was $\frac{m_r}{n}$.

Suppose, finally, that we are told that N drawings have been made and that on M of these occasions a white counter has been drawn. The question is this. Given this statistical information, the rule according to which the experimenter chose the bag from which he made all his subsequent drawings, the information detailed above about the constitution of the contents of the various bags, and the rule according to which the drawings were made when the bag to be used had been chosen, what is the probability that the drawings were made from a particular bag B_r? Or, what is equivalent, what is the probability that the probability of drawing a white counter on each occasion throughout the experiment was $\frac{m_r}{n}$?

When this problem is solved we can raise the following question. Can we say that there is *one particular bag* (and, if so, which one) that is more likely than any other to have been the one on which the experiments were performed, in view of the data enumerated above? It will be found that it is possible, on certain assumptions, to give a certain answer to this question. This answer, subject to these conditions, constitutes the *Inverse Principle of Greatest Probability*.

We must now generalise the problem, and then solve it. Suppose that a series of N trials of instances of Q has been made under Bernoullian conditions, which we will denote by h, and that the proportion of R's among them has been p. Suppose

it is known that $\mathrm{R}(x)/\mathrm{Q}(x)\ \&\ h$ must have had one or other of the values p_1, or p_2, or . . . p_z, but it is not known which one of them it had. Suppose further that the subject of the experiment was selected in accordance with a certain rule k, such that relative to k the probability that $\mathrm{R}(x)/\mathrm{Q}(x)\ \&\ h = p_1$ is q_1, the probability that it equals p_2 is q_2, and so on for the other alternatives. What is the probability, with respect to all these data, that the experiment was done on a subject for which

$$\mathrm{R}(x)/\mathrm{Q}(x)\ \&\ h = p_r\ ?$$

Our problem, then, is to evaluate the expression

$$\mathrm{R}(x)/\mathrm{Q}(x)\ \&\ h = p_r\ .\ /\ :\ k\ .\ \&\ .\ f_{\mathrm{N}}(\mathrm{R}\ ;\ \mathrm{Q}) = p.$$

In order to do this we will introduce the following temporary abbreviations. We will write the single letter P for the proposition $f_{\mathrm{N}}(\mathrm{R}\ ;\ \mathrm{Q}) = p$. We will write the single letter P_r for the proposition $\mathrm{R}(x)/\mathrm{Q}(x)\ \&\ h = p_r$. The expression to be evaluated then can be written $\mathrm{P}_r/k\ \&\ \mathrm{P}$.

Now by Lemma VII (*Bayes's Principle*) this is equal to

$$\frac{\mathrm{P}_r/k \times \mathrm{P}/\mathrm{P}_r\ \&\ k}{\sum_{r=1}^{r=z} \mathrm{P}_r/k \times \mathrm{P}/\mathrm{P}_r\ \&\ k}$$

which is equal to

$$\frac{q_r \times \mathrm{P}/\mathrm{P}_r\ \&\ k}{\sum_{r=1}^{r=z} q_r \times \mathrm{P}/\mathrm{P}_r\ \&\ k}.$$

But such expressions as $\mathrm{P}/\mathrm{P}_r\ \&\ k$ have already been evaluated in the first part of the Direct Bernoulli Theorem. For, when we once more write them out in full, they are of the form

$$f_{\mathrm{N}}(\mathrm{R}\ ;\ \mathrm{Q}) = p\ .\ /\ :\ k\ .\ \&\ .\ \mathrm{R}x/\mathrm{Q}x\ \&\ h = p_r.$$

So this part of the problem is now solved in principle, though it cannot be solved in detail unless the rule k enables us to give determinate values to q_1, q_2, . . . q_z, *i.e.*, to the antecedent probabilities that such and such a subject has been chosen for the experiment.

It remains to consider for what value of P_r the probability $\mathrm{P}_r/k\ \&\ \mathrm{P}$ is a maximum. No *general* answer can be given to this question. Suppose, however, that the rule k is such that $q_1 = q_2 = \ldots\ q_z$. Then the expression given above for $\mathrm{P}_r/k\ \&\ \mathrm{P}$ reduces to

$$\frac{\mathrm{P}/\mathrm{P}_r\ \&\ k}{\sum_{r=1}^{r=z} \mathrm{P}/\mathrm{P}_r\ \&\ k}.$$

Now this will be a maximum when its numerator $P/P_r \,\&\, k$ is a maximum. But $P/P_r \,\&\, k$ is an abbreviated way of writing

$$f_N(R\,;\,Q) = p_r \,./:\, k \,.\,\&\,.\, R(x)/Q(x)\;\&\;h = p\,;$$

and we have already proved in the *Direct Principle of Greatest Probability* that this is a maximum when $p_r = p$.

We can now state the *Inverse Principle of Greatest Probability*, which is the proposition that we have just proved. It runs as follows. Suppose that a series of N trials of instances of Q has been made under the Bernoullian conditions, and that a proportion p of these Q's have been found to be R's. Suppose further that the probability of a Q being R might have had any of the z values $p_1, p_2, \ldots p_z$, and that relative to the information supplied it was *equally likely to have any of these values*. Then, relative to all the suppositions detailed above, the most probable value of the probability of a Q being an R in this experiment is that one of the possible values $p_1, \ldots p_z$ which is nearest to p.

Before leaving this part of the Inverse Bernoulli Theorem it will be worth while to revert for a moment to our original example of the various bags from one of which the drawings are to be made. We shall then be able to see the point of the condition that all the alternative possible values of $R(x)/Q(x)\;\&\;h$ must be equally probable with respect to the rule k if the Inverse Principle of Greatest Probability is to hold. Suppose that this condition was not fulfilled. Suppose, *e.g.*, that there were eleven bags, one with 0 %, one with 10 %, one with 20 %, . . . and one with 100 % of white counters in it; and suppose that, instead of it being equally likely that any one of these bags would be the one chosen as the subject of the experiment, the rule was such that it was very much more likely that the experiment would have been performed with the bag containing 20 % of whites than with any of the others. Suppose that 100 trials were made and that in 48 % of them the counter drawn was white. Then it would obviously be quite unjustifiable to conclude in accordance with the Inverse Principle of Greatest Probability that it is most probable that the experiment was done on the bag containing 50 % of whites. For we should be faced with two alternative improbabilities. (i) The bag that was most likely to have been chosen according to the rule of selection, *viz.*, that containing 20 % of white counters, may in fact have been chosen. But, if so, the 100 trials made on it have resulted in a most improbably large proportion of white drawings. Or (ii) the bag chosen may have been the one which

contains 50 % of white counters. If so, the 48 % of white drawings which were obtained is a very probable result. But, on the other hand, the selection of this bag was a very improbable event. Now the *final* probability that the drawings were made from such and such a bag depends jointly on the *initial* probability that that bag would have been chosen and on the probability that, if it were chosen, the results actually obtained would have followed. These two factors point in opposite directions in the case supposed, and so it is not surprising that no general conclusion can be drawn.

(2·2) *Inverse Principle of Great Numbers.*—We will begin, as usual, with a concrete example. Suppose now that there are $n + 1$ bags, each containing n counters. We will call them B_0, B_1, . . . B_n. The number of white counters is to be 0 in B_0, 1 in B_1, . . ., and n in B_n. Then suppose n to be increased without limit. In that case every proper fraction between 0/1 and 1/1, together with these two end-points, will be represented by one and only one of the bags. We know that *one* of these bags has been selected to be the subject of the experiment, but we do not know *which* one. The selection has been made in accordance with a certain rule k with respect to which the probability that such and such a bag has been selected is so-and-so. N trials are made with the selected bag, whatever it may be, and pN of these have been white.

The proposition to be proved may be stated as follows. Let δ and ϵ be any two pre-assigned quantities. Then (*subject to one condition about k, which will be stated below*), no matter how small δ and ϵ may be, there is a number ν of trials such that, if N exceeds ν, the probability that the bag selected was one of those in which the proportion of whites does not differ from p by more than δ does not differ from 1 by more than ϵ.

This proposition may be stated more colloquially as follows. By making the series of trials long enough you can always ensure that the probability of the bag selected being one in which the proportion of white counters differed by as little as you please from the proportion of white drawings in the series of trials differs by as little as you please from 1.

The condition required is as follows. The rule k, in accordance with which the bag on which the trials are to be made is selected, may be of any kind, *provided only that the probability with respect to k of a bag with the proportion p of white counters being selected is not* 0.

It remains to generalise this proposition and to prove it.

The general statement is as follows. Suppose (i) that an

experiment has been done under Bernoullian conditions, and that N instances of Q have been observed and that pN of them have been found to be R. (ii) That the value of $R(x)/Q(x)$ & h may have been anything from 0 to 1, both inclusive, and it is not known which of these possible values it had. (iii) That the probability that the value of $R(x)/Q(x)$ & h was in the immediate neighbourhood of p is not 0 with respect to k, where k is the only relevant datum available to us about the value of $R(x)/Q(x)$ & h. Let δ and ϵ be any pre-assigned quantities. Then, however small δ and ϵ may be, there is a number ν such that if N exceeds ν the probability, with respect to the above three supposals, that $R(x)/Q(x)$ & h did not differ from p by more than δ does not differ from 1 by more than ϵ. The accurate expression for this in our symbolism is

$$(\delta, \epsilon) : : : (\exists \nu) : : . N > \nu \supset_N : : R(x)/Q(x) \,\&\, h \doteq p \pm \delta \,./ : .$$
$$k : \& : f_N(R ; Q) = p : \& : R(x)/Q(x) \,\&\, h = p \,./k \neq 0 : . \geqslant 1 - \epsilon.$$

The proof of this proposition may be stated as follows. Let us divide the interval between 0 and 1 into a very large number μ of very small sub-intervals of equal length η adjoined to each other, so that $\mu\eta = 1$. Then the proposition 'The value of $R(x)/Q(x)$ & h lies somewhere between 0 and 1' is a disjunction of μ mutually exclusive alternatives of the form 'The value of $R(x)/Q(x)$ & h lies between 0 and η, or between η and 2η, or . . . between $1 - \eta$ and 1'. Let us denote a typical one of these alternatives, *viz.*, 'The value of $R(x)/Q(x)$ & h lies between $r\eta$ and $(r + 1)\eta$' by P_r. It is evident that r will range between 0 and $\mu - 1$.

In precisely the same way the proposition 'The value of $R(x)/Q(x)$ & h lies somewhere between $p - \delta$ and $p + \delta$', is a disjunction of $2\mu\delta$ mutually exclusive alternatives P_r, where r ranges from $(p - \delta)\mu$ to $(p + \delta)\mu - 1$.

Let us denote the statistical proposition $f_N(R ; Q) = p$ by P, as we did before. Then the question which concerns us is: To what limiting value, if any, does the probability

$$P_{(p-\delta)\mu} \vee \ldots\ldots P_{(p+\delta)\mu-1} : /P \,\&\, k$$

approach as N is made greater and greater?

It is evident that we can apply Lemma VIII (*The Extended Bayes Principle*) to the probability in question. It is therefore equal to

$$\frac{\overset{r=(p+\delta)\mu-1}{\underset{r=(p-\delta)\mu}{S}} P_r/k \times P/P_r \,\&\, k}{\overset{r=\mu-1}{\underset{r=0}{S}} P_r/k \times P/P_r \,\&\, k}.$$

Now consider the denominator of this fraction. Plainly it can be regarded as the sum of the following three parts, *viz.*, (i) the terms from $r = 0$ to $r = (p - \delta)\mu - 1$, both inclusive; (ii) the terms which also occur in the numerator; and (iii) the terms from $r = (p + \delta)\mu$ to $r = \mu - 1$, both inclusive. Let us call the first and the third of these the 'fringes', and the second of them the 'kernel'. The fraction is, then, of the form

$$\frac{B}{A + B + C}.$$

If we can show that, as N is increased without limit, (a) the fringes differ by as little as we please from 0, whilst (b) the kernel remains finite, we shall have shown that the fraction differs by as little as we please from 1 if N be made great enough. And this is what we have to prove. (The second condition is needed as well as the first; for, if the first were fulfilled without the second, the fraction would assume the indeterminate form 0/0.)

The argument may be put as follows. The symbol $P/P_r \,\&\, k$ is simply an abbreviation for

$$f_N(R\,;\,Q) = p\,./:k\,.\,\&\,.\,R(x)/Q(x)\,\&\,h \backsimeq (r + \tfrac{1}{2})\eta \pm \tfrac{1}{2}\eta.$$

Now, according to the *Direct Principle of Great Numbers*, as N increases without limit this will differ by as little as we please from 1 when $r\eta = p$, *i.e.*, when $r = \dfrac{p}{\eta}$, *i.e.*, when $r = \mu p$. Therefore by increasing N sufficiently we can make every term of the form $P/P_r \,\&\, k$ for which r is *not* in the immediate neighbourhood of μp differ by as little as we please from 0. Therefore, however small δ may be, if N be made large enough each term in the fringes will consist of a term of the form P_r/k multiplied by a term which is vanishingly small. Now all terms of the form P_r/k are positive and not greater than 1 by Postulate (ii). Therefore, whatever their actual values may be, the fringes can be made to differ by as little as we please from 0 by sufficiently increasing N. On the other hand, provided that the terms of the form P_r/k are not equal to 0 in the immediate neighbourhood of $r = \mu p$, the kernel will not vanish. And so the fraction will approach as nearly as we please to 1, no matter what may be the values of P_r/k for the various values of r, provided only that P_r/k is not zero in the immediate neighbourhood of $r = \mu p$. And this is what we set out to prove.

The following three remarks are worth making before we leave this theorem. (1) The argument just ended shows the precise force of the condition that $R(x)/Q(x) \,\&\, h = p\,./k$ must not be zero. This condition is needed in order that the fraction may not reduce to the indeterminate form 0/0.

(2) If we go back to the example of the bags we shall see that the conclusion, and the conditions under which it is reached, are in accord with common sense. Provided that there is a finite antecedent probability that the bag which was used for the experiment contained a proportion of white counters in the immediate neighbourhood of p, it does not matter what may be the antecedent probabilities of the bag containing other proportions of white counters. For suppose that the antecedent probability of the bag containing a proportion of white counters widely different from p is quite high. Nevertheless, as more and more trials were made with it and the proportion of white drawings still remained obstinately in the neighbourhood of p, it would become more and more unlikely that with such a bag such results would be obtained. Eventually the improbability that the actual results should be obtained from a bag of the kind which is antecedently most likely to have been used would (to quote a happy phrase of Dr. Harold Jeffreys) 'swamp' the antecedent probability that the bag was of the kind supposed.

(3) The following considerations might strike an attentive reader as paradoxical. The Inverse Principle of Great Numbers seems to be a much more determinate and exciting proposition than the Inverse Principle of Greatest Probability. Yet in order to deduce the former we had to impose a much more rigid condition on the possible values of $\mathrm{R}(x)/\mathrm{Q}(x)$ & $h = p_r \,./k$ than was needed in deducing the latter. I think that this apparent paradox is removed when we remember the following fact. The Inverse Principle of Great Numbers is itself a conditional proposition, and, although its *consequent* (that a certain probability will differ as little as we please from 1) is highly determinate and exciting, this is subject to a very severe *antecedent* condition, *viz.*, 'if the frequency-ratio of R's among Q's is p when N is indefinitely increased'. The Inverse Principle of Greatest Probability does not contain any such limitations within itself, and therefore it is not surprising that more rigid conditions have to be inserted in the premisses from which it is proved.

Theorem 3. *The Statistical Generalisation Theorem.*—In terms of counters and bags the problem with which we are now concerned may be stated as follows. Suppose that N drawings have been made under Bernoullian conditions from a bag of unknown constitution, and that pN of them have been white. Suppose that the process of drawing and replacing were now to be continued indefinitely. Then (1) what is the most likely proportion of white drawings, relative to this information, in the indefinitely prolonged series? And (2) as N is indefinitely

increased does the probability that the proportion of white drawings in the indefinitely prolonged series will be that which is most probable approach indefinitely nearly to 1 ?

The general line of argument is obvious at once. It will be in two stages; the first backwards from the observed results to the probable constitution of the bag from which the counters have been drawn, and the second forwards from this to the probable results of future drawings. The backward step will use the Inverse Bernoulli Theorem, and the forward step will use the Direct Bernoulli Theorem. According to the Inverse Principle of Greatest Probability, if the antecedent probability of the bag being of any one constitution is the same as that of its being of any other, the most probable proportion of white counters in the bag in view of the observed results is p. According to the Direct Principle of Greatest Probability, if the proportion of white counters in the bag used is p the most probable proportion of white drawings in a series of drawings from it is p. In this way one can see in outline that it is at any rate plausible to conclude that the most probable proportion of white drawings in the indefinitely extended series will be the same as the proportion in the finite series of actual drawings, provided only that the antecedent probability of the bag containing any of the logically possible proportions of white counters is the same.

A similar rough argument can be used in connexion with the second part of the problem. By the Inverse Principle of Great Numbers, if N be made great enough and the proportion of white drawings in the actual series be p, the probability that the proportion of white counters in the bag used is p differs as little as we please from 1, provided only that the antecedent probability of the bag having this constitution is not zero. By the Direct Principle of Great Numbers, if the proportion of white counters in the bag be p and the number of drawings be made great enough, then the probability that the proportion of white drawings will differ from p by as little as we please will differ from 1 by as little as we please. So we can see in outline that it is plausible to conclude that, if the series of actual observations is long enough, the probability that the proportion of white drawings in the indefinitely extended series of possible drawings is the same as the proportion in the actual series will differ by as little as we please from 1. The condition here is that the antecedent probability that the bag used was one that contained this proportion of white counters shall not be zero. (It is important to notice here that we have two series of drawings to consider,

(i) a very long but finite series of actual drawings, and (ii) an indefinitely long extension of possible further drawings. The former is involved in the supposal, and the latter in the proposal.)

We must now generalise the Theorem and try to provide a proper proof, as distinct from a sketchy outline of an argument, for it.

I shall symbolise the proposition 'The proportion of R's in an indefinitely extended series of Q's is p' by $f_\infty(\mathrm{R}\,;\,\mathrm{Q}) = p$. It will be remembered that we discussed the definition of this proposition under the heading *Statistical Propositions* in the first Section of this paper (Part I). We saw that Hr. von Wright gave a rather complicated definition. But this definition is logically equivalent to the following, which was there labelled (iii), *viz.*,

$$(\delta) : (\exists\nu) \,.\, N > \nu \supset_{\mathrm{N}} f_{\mathrm{N}}(\mathrm{R}\,;\,\mathrm{Q}) \doteq p \pm \delta.$$

And this is simply the statement that $f_{\mathrm{N}}(\mathrm{R}\,;\,\mathrm{Q})$ approaches a limit as N is indefinitely increased, and that that limit is p. So we may write

$$f_\infty(\mathrm{R}\,;\,\mathrm{Q}) = p \,.\, \equiv \,.\, \operatorname*{Lt}_{\mathrm{N}=\infty} f_{\mathrm{N}}(\mathrm{R}\,;\,\mathrm{Q}) = p.$$

This being understood, the two propositions which we have to prove may be stated as follows:

(3·1) *Statistical Principle of Greatest Probability.*—If

$$\mathrm{R}(x)/\mathrm{Q}(x)\ \&\ h = y \,.\, /k$$

has the same value for all values of y from 0 to 1, both inclusive, then $f_\infty(\mathrm{R}\,;\,\mathrm{Q}) = z \,.\, /:k \,.\, \&\,.\, f_{\mathrm{N}}(\mathrm{R}\,;\,\mathrm{Q}) = p$ is a maximum when $z = p$.

(3·2) *Statistical Principle of Great Numbers.*—Provided that $\mathrm{R}(x)/\mathrm{Q}(x)\ \&\ h = p \,.\, /k$ is not equal to 0, then

$$(\delta, \epsilon) :: . \,(\exists\nu) :: N > \nu \supset_{\mathrm{N}} : . f_\infty(\mathrm{R}\,;\,\mathrm{Q}) \doteq p \pm \delta \,.\, / : f_{\mathrm{N}}(\mathrm{R}\,;\,\mathrm{Q}) = p \,.\, \&\,.\, k : \geqslant 1 - \epsilon.$$

Now Hr. von Wright claims to prove these propositions in a way which is certainly very neat and interesting if it is valid. I cannot say that I feel altogether happy about these proofs, especially that of the Statistical Principle of Greatest Probability. The argument depends on (i) a certain general proposition in formal probability which is asserted but not proved, and (ii) the assertion that the Direct and the Inverse Bernoulli Theorems are equivalent respectively to a certain pair of other propositions. These propositions are stated in words and not in symbols, and the equivalence is not explicitly proved. The general proposition which forms the first of these two premisses is certainly true, and I shall prove it as a Lemma. I shall then state formally the pair of propositions which are alleged to be

equivalent respectively to the Direct and the Inverse Bernoulli Theorems. Finally I shall complete the argument as stated by Hr. von Wright. The reader should then be in a position to judge for himself whether it is conclusive or not. We begin then with what I will call *Lemma IX.*

Lemma IX. If a and b be any two propositions, and

$$a/b \,\&\, h = b/a \,\&\, h = 1,$$

then if c be any third proposition $a/c \,\&\, h = b/c \,\&\, h$.

The proof is as follows.

By Lemma VI $\quad a/b \,\&\, h = \dfrac{a/h \times b/a \,\&\, h}{b/h}.$

But by hypothesis $a/b \,\&\, h = b/a \,\&\, h = 1$. Therefore

$$a/h = b/h. \quad . \quad . \quad . \quad . \quad (1)$$

Now $c \equiv : c \,\&\, b \,.\, \vee \,.\, c \,\&\, \bar{b}$.

Therefore, by Lemma IV,

$$c/a \,\&\, h = b/a \,\&\, h \times c/b \,\&\, a \,\&\, h + \bar{b}/a \,\&\, h \times c/\bar{b} \,\&\, a \,\&\, h.$$

But $b/a \,\&\, h = 1$ by Hypothesis, and therefore $\bar{b}/a \,\&\, h = 0$ by Lemma I.

Therefore $\quad c/a \,\&\, h = c/b \,\&\, a \,\&\, h. \quad . \quad . \quad . \quad (2{\cdot}1)$

Again $\quad c \equiv : c \,\&\, a \,.\, \vee \,.\, c \,\&\, \bar{a}.$

From this and the fact that $a/b \,\&\, h = 1$ by hypothesis we can prove in exactly the same way that

$$c/b \,\&\, h = c/a \,\&\, b \,\&\, h. \quad . \quad . \quad . \quad (2{\cdot}2)$$

Therefore $\quad c/a \,\&\, h = c/b \,\&\, h. \quad . \quad . \quad . \quad (3)$

Now by Lemma IV.

$$c/a \,\&\, h = \frac{c/h \times a/c \,\&\, h}{a/h}. \quad . \quad . \quad (4{\cdot}1)$$

and

$$c/b \,\&\, h = \frac{c/h \times b/c \,\&\, h}{b/h}. \quad . \quad . \quad . \quad (4{\cdot}2)$$

Therefore from (3) $\quad \dfrac{a/c \,\&\, h}{a/h} = \dfrac{b/c \,\&\, h}{b/h}.$

But from (1) $a/h = b/h$.

Therefore $\quad a/c \,\&\, h = b/c \,\&\, h. \qquad$ Q.E.D.

Now Hr. von Wright asserts, in effect, that (i) the Direct Principle of Great Numbers is equivalent to

$$f_\infty(\mathrm{R} \,;\, \mathrm{Q}) = p \,.\, / : k \,.\, \&\, .\, \mathrm{R}(x)/\mathrm{Q}(x) \,\&\, h = p : = 1 \quad . \quad (\alpha)$$

and that (ii) the Inverse Principle of Great Numbers is equivalent to

$$\mathrm{R}(x)/\mathrm{Q}(x) \,\&\, h = p \,.\, / : k \,.\, \&\, .\, f_\infty(\mathrm{R} \,;\, \mathrm{Q}) = p : = 1 \qquad (\beta)$$

Let us accept this for the sake of argument. Then we can apply Lemma IX. In Lemma IX for a put $\mathrm{R}(x)/\mathrm{Q}(x)\ \&\ h = p$; for b put $f_\infty(\mathrm{R}\,;\,\mathrm{Q}) = p$; and for c put $f_\mathrm{N}(\mathrm{R}\,;\,\mathrm{Q}) = p$. Then $a/b\ \&\ h = b/a\ \&\ h = 1$. And therefore $a/c\ \&\ h = b/c\ \&\ h$. That is

$$f_\infty(\mathrm{R}\,;\,\mathrm{Q}) = p\,./\,:k\,.\,\&\,.f_\mathrm{N}(\mathrm{R}\,;\,\mathrm{Q}) = p :=: \mathrm{R}(x)/\mathrm{Q}(x)\ \&\ h = p\,./\,:k\,.\,\&\,.f_\mathrm{N}(\mathrm{R}\,;\,\mathrm{Q}) = p \quad \text{(A)}$$

Having established Equation (A) in this way, Hr. von Wright uses it to prove the Statistical Principle of Greatest Probability and the Statistical Principle of Great Numbers as follows.

(i) Subject to the condition that $\mathrm{R}(x)/\mathrm{Q}(x)\ \&\ h = z\,./k$ has the same value for all values of z from 0 to 1, both inclusive, the Inverse Principle of Greatest Probability tells us that $\mathrm{R}(x)/\mathrm{Q}(x)\ \&\ h = z\,./\,:k\,.\,\&\,.f_\mathrm{N}(\mathrm{R}\,;\,\mathrm{Q}) = p$ is a maximum when $z = p$. But when $z = p$ this expression is equal to

$$f_\infty(\mathrm{R}\,;\,\mathrm{Q}) = p\,./\,:k\,.\,\&\,.f_\mathrm{N}(\mathrm{R}\,;\,\mathrm{Q}) = p,$$

by Equation (A). Therefore, subject to the condition mentioned above,

$$f_\infty(\mathrm{R}\,;\,\mathrm{Q}) = z\,./\,:k\,.\,\&\,.f_\mathrm{N}(\mathrm{R}\,;\,\mathrm{Q}) = p$$

is a maximum when $z = p$. And this is the Statistical Principle of Greatest Probability.

(ii) Subject to the condition that $\mathrm{R}(x)/\mathrm{Q}(x)\ \&\ h = p\,./k$ is not zero, the Inverse Principle of Great Numbers tells us that the expression on the right-hand side of Equation (A) approaches as near as we please to 1 if N be sufficiently increased. Therefore the same is true of the expression on the left of Equation (A). And this is the Statistical Principle of Great Numbers.

I have no doubt that this is a fair and accurate account, in our symbolism, of Hr. von Wright's arguments. Are they valid? We must distinguish two questions: (1) Is the proof of Equation (A) valid? (2) If so, is the reasoning from it valid in (i), or in (ii), or in both? I will take these two questions in turn.

(1) Equation (A) certainly follows from the premisses; and one of the premisses, *viz.*, Lemma IX, is certainly true. So the only possible doubt about the proof of it is a doubt whether Proposition (α) is a legitimate transformation of the Direct Principle, and Proposition (β) of the Inverse Principle, of Great Numbers.

Now when I put these propositions into words and consider their meaning it does seem to me that these alleged equivalences

are highly plausible. But this is not enough. It ought to be possible, by using the definition of $f_\infty(\mathrm{R}\,;\mathrm{Q}) = p$ given above, to prove that the Direct Principle of Great Numbers is equivalent to (α) and that the Inverse Principle of Great Numbers is equivalent to (β). I must confess that I have not succeeded in doing this to my own satisfaction. It does not follow that it cannot be done, but my failure to do it prevents me from feeling altogether comfortable about the proof of Equation (A). I would remark that it would be enough for Hr. von Wright's purpose to show that the Direct Principle entails (α) and that the Inverse Principle entails (β). It is not necessary that these entailments should be reversible. But I have failed to prove even these milder propositions to my own satisfaction, though I see no reason to doubt that they are true.

(2) Suppose that (α) and (β) can be shown to follow from the Direct and the Inverse Principles of Great Numbers respectively, and that Equation (A) can therefore be proved to be true. Then I can see nothing to criticise in the deduction from it of the Statistical Principle of *Great Numbers*. But I do not feel comfortable about the attempted deduction from it of the Statistical Principle of *Greatest Probability*. My difficulty here is the following. Suppose that Equation (A) had been the proposition: '*For every value of* z

$$f_\infty(\mathrm{R}\,;\mathrm{Q}) = z\,./:k\,.\,\&\,.f_\mathrm{N}(\mathrm{R}\,;\mathrm{Q}) = p :$$
$$= :\mathrm{R}(x)/\mathrm{Q}(x)\,\&\,h = z\,./:k\,.\,\&\,.f_\mathrm{N}(\mathrm{R}\,;\mathrm{Q}) = p\,'.$$

Then no doubt it would follow that, since the right-hand side is a maximum for $z = p$, the left-hand side is also a maximum for $z = p$. But this is not what Equation (A) asserts. It asserts the equality between the left-hand and the right-hand side *only* for the particular case of $z = p$. This seems to me to make the argument a *non-sequitur*.

I shall now suggest an alternative way of proving the two parts of the Statistical Generalisation Theorem.

Let us, as before, divide the interval between 0 and 1 into a very large number μ of very small adjoined sub-intervals each of length η, so that $\mu\eta = 1$. As before, let p_r stand for the proposition '$\mathrm{R}(x)/\mathrm{Q}(x)\,\&\,h$ lies between $r\eta$ and $(r+1)\eta$.' Let us write $\mathrm{P}_\infty(z)$ for the proposition $f_\infty(\mathrm{R}\,;\mathrm{Q}) = z$; and let us write $\mathrm{P}_\mathrm{N}(p)$ for the proposition $f_\mathrm{N}(\mathrm{R}\,;\mathrm{Q}) = p$. Then the two propositions which we have to prove may be written:

(i) $\mathrm{P}_\infty(z)/\mathrm{P}_\mathrm{N}(p)\,\&\,k$ is a maximum when $z = p$, provided that P_r/k has the same value for all values of r; and

(ii) $\underset{\mathrm{N}\to\infty}{\mathrm{Lt}}\ \mathrm{P}_\infty(p)/\mathrm{P}_\mathrm{N}(p)\,\&\,k = 1$, provided that $\mathrm{P}_{\mu p}/k$ is not zero.

The proofs are as follows.

(i) Since the alternatives P_0 to $P_{\mu-1}$ are mutually exclusive and collectively exhaustive, we have, by Lemma V,

$$P_\infty(z)/P_N(p) \,\&\, k = \sum_{r=0}^{r=p-1} P_r/P_N(p) \,\&\, k \times P_\infty(z)/P_r \,\&\, P_N(p) \,\&\, k \quad (1)$$

We note also that *Proposition* (α) can be written in the present notation as

$$P_\infty(z)/P_{\mu z} \,\&\, k = 1.$$

Again, the *Inverse Principle of Greatest Probability* can be written in our present notation as: '$P_{\mu z}/P_N(p) \,\&\, k$ is a maximum when $z = p$, provided that P_r/k has the same value for all values of r'.

Consider the typical factor $P_\infty(z)/P_r \,\&\, P_N(p) \,\&\, k$ in the expression on the right-hand side of Equation (1). The term $P_N(p)$ is superfluous, when conjoined in the supposal with P_r, as regards the proposal $P_\infty(z)$. For the fact that, in a series of N trials of Q's under Bernoullian conditions, such and such a proportion have been found to be R is relevant only in the follow ing way to the probability that the proportion of R's in an indefinitely extended series of trials will be so-and-so. It is relevant only in so far as it affects the probability that $R(x)/Q(x) \,\&\, h$ had such and such a value throughout the experiment. But the datum P_r settles this question *independently;* for it tells us *with certainty* that the value of $R(x)/Q(x) \,\&\, h$ lay between $r\eta$ and $(r+1)\eta$. Therefore $P_N(p)$, when conjoined in the supposal with P_r, is irrelevant to the proposal $P_\infty(z)$. So we can replace each such factor as $P_\infty(z)/P_r \,\&\, P_N(p) \,\&\, k$ on the right-hand side of Equation (1) by the simpler factor $P_\infty(z)/P_r \,\&\, k$.

Now, by Proposition (α), $P_\infty(z)/P_r \,\&\, k$ is equal to 1 when $r = \mu z$, and is equal to 0 for all other values of r. So the expression on the right-hand side of Equation (1) reduces to the single term $P_{\mu z}/P_N(p) \,\&\, k$. And, by the *Inverse Principle of Greatest Probability,* this is a maximum when $z = p$, provided that P_r/k has the same value for all values of r. Therefore the same is true of the left-hand side of Equation (1); *i.e.*, $P_\infty(z)/P_N(p) \,\&\, k$ is a maximum when $z = p$, provided that P_r/k has the same value for all values of r. Q.E.D.

(ii) By precisely the same reasoning as we used above we can show that

$$P_\infty(p)/P_N(p) \,\&\, k = P_{\mu p}/P_N(p) \,\&\, k. \quad . \quad . \quad (2)$$

But the *Inverse Principle of Great Numbers,* written in our present notation, simply is the proposition that $\operatorname{Lt}_{N\to\infty} P_{\mu p}/P_N(p) \,\&\, k = 1$

provided that P_{μ_p}/k is not zero. Therefore, provided that P_{μ_p}/k is not zero, $\underset{N\to\infty}{\text{Lt}}\ P_\infty(p)/P_N(p)\ \&\ k = 1$. Q.E.D.

The following points should be noted about these proofs. (1) Like Hr. von Wright's, they presuppose that Proposition (α) is guaranteed by the Direct Principle of Great Numbers. (2) Unlike his, they do not presuppose Proposition (β) and they make no use of Lemma IX. (3) If, as it seems to me, there is a defect in the argument by which Hr. von Wright claims to deduce the Statistical Principle of Greatest Probability from Equation (A), this defect is avoided.

There is one further matter to be mentioned before we leave Theorem 3. It will be remembered that the Statistical Principle of Greatest Probability has been proved only subject to the condition that p_r/k has the same value for all values of r, *i.e.*, that all possible values from 0 to 1 of $R(x)/Q(x)\ \&\ h$ are equally likely with respect to k. Now Hr. von Wright claims that this is entailed by the following condition, *viz.*, that the datum $f_N(R\,;\,Q) = p$ (*i.e.*, $P_N(p)$) is the *only* information which we are given that is relevant to the proposal $f_\infty(R\,;\,Q) = p$ (*i.e.*, to $P_\infty(p)$). So he takes this latter assumption as the condition under which the Statistical Principle of Greatest Probability holds good.

His argument, which seems to me to be quite sound, is as follows.

By Lemma V

$$P_\infty(z)/k = \overset{r=\mu-1}{\underset{r=0}{S}} P_r/k \times P_\infty(z)/P_r\ \&\ k.$$

Now suppose that P_r/k does *not* have the same value for all values of r, but is a maximum for a certain particular value of r, *e.g.*, when $r = \mu\lambda$. By the Direct Principle of Greatest Probability the factor $P_\infty(z)/P_{\mu\lambda}\ \&\ k$ will be greatest if $z = \lambda$. Therefore, if and only if $z = \lambda$, the greatest of the factors P_r/k in the series on the right (*viz.*, $P_{\mu\lambda}/k$) will be multiplied by the greatest possible value of the factor associated with it (*viz.*, $P(\lambda)/P_{\mu\lambda}\ \&\ k$). Therefore, on our supposition, k will be more favourably relevant to the proposal $P_\infty(\lambda)$ than to any similar proposal in which the value of z is other than λ. That is to say $P_N(p)$ will *not* be the *only* information which we are given that is relevant to the proposal $P_\infty(p)$. By contraposing this entailment we reach Hr. von Wright's conclusion.

(*To be concluded.*)