St John's College
Cambridge.

March 1. 1934

Dear Fisher,

Thanks for your letter. I think that even if we can't agree and have to leave a lot for other people to argue about, we ought to be able to agree about what our points of difference are. At present it seems that the relations between us are of the form 'A thinks that B has done all the things that B has been at special trouble to avoid' ; where if A = J, B = F and conversely. With me might I think be included F.P.Ramsey and C.D.Broad; there are two very important papers by the latter in Mind vols 27 and 29, which I read when they came out, and they seem (especially the first) to have influenced me to the extent that I assimilated the ideas and forget where they came from. I'll have to make a belated acknowledgment somewhere.

On the question of ratios of infinites first. In your Phil.Trans. you quite explicitly excluded the notion of a limit and stated the definition of a probability as the ratio of ~~xxx~~ two infinite numbers. The limit was Venn's dodge, and you dropped it. But when it comes to the point neither you nor Venn use your own definitions. You have never taken an infinite class and counted the number of a sub-class within it, nor has Venn ever found a probability as the limit of a ratio when the number of trials tends to infinity. (By the way you use the word 'probability' in your definition ; but later on you generally use 'frequency'. I suppose you mean the same thing by both, but am not sure.) When you want actual numerical estimates you assess directly the frequency of a sample in the form ${}^{r}C_{l}\,{}^{n-r}C_{n-l}/{}^{n}C_{r}$. So far as I can see this is got by counting cases in a perfectly correct way and then saying that all cases are equally probable. On the a priori view of probability there is no more to be said ; on the frequency view it may be right or not, but the point I want to make is that you don't avoid making an <u>a priori</u> assumption. You have to assume that if you take an indefinitely large number of classes of number n, each of which contains r $\phi$'s and sample them, the ratios of compositions of the samples will occur in just the frequency given by your estimated probability. You do not know this by experience and therefore it is an <u>a priori</u> postulate. You may think that it is more plausible than the direct use of non-sufficient reason ; but that is not a reason for condemning the latter on the ground that it is not known by experience. So far as I can see your practical results* can be taken over into the <u>a priori</u> theory of probability without change of their quantitative statement, though some of them may need a change of language.

*In the Phil. Trans. memoir; I have not followed your later work in much detail.

To take another case ; suppose that by your methods you compare two methods of growing potatoes and show that in 90 per cent of cases method A will give a greater yield than method B. Suppose a farmer asks you 'What reason is there to suppose that <u>I</u> will get a greater yield by method A?' It seems to me that your only answer is in terms of the view that probability is intelligible without definition.

There is a criticism of the <u>a priori</u> view that I often meet and have never succeeded in understanding, and you give me some clue to it at last. It is the fundamental objection to the idea that $P(p|q)$ has a definite value whatever p and q may be. I am getting inclined to think that behind this objection is an idea that I think that q is <u>relevant</u> to the truth of p in all cases, which is another matter altogether. In the majority of cases we have things like $P(p \mid qr) = P(p|q)$ whatever r may be over a very wide range, i.e., in words, r is irrelevant to the truth of p given q. E.g., on my present knowledge of your movements it is as likely as not that you will be in Cambridge before you are at Rothamsted ; the probability is $\frac{1}{2}$ . If I find that you still have a house at Rothamsted the probability sinks nearly to 0 ; but it is only such quite special additions to information that will materially alter the probability.

You also seem to suggest in your paper that when I assess a prior probability I am expressing an opinion on the ratio of the numbers of cases in the world. My attitude, on the contrary, is that it is only on my view that it is possible to ~~avoid~~ make any progress without expressing such an opinion. I started from the pure phenomenalist position, but found that phenomenalism needs a good deal of amplification before it can deal with the problem of inference. We don't know all about the world to start with ; our knowledge by experience consists simply of a rather scattered lot of sensations, and we cannot get any further without some a priori postulates. My problem is to get these stated as clearly as possible. The tests available cannot be experimental; the conditions required, in my view, are that we want just enough a priori hypothesis to make it possible to settle the rest by experience. We cannot exclude logically the possibility that when a law looks well established the next attempt at verification may jigger it up altogether, so that general laws are never certain at any stage but only have a certain degree of probability I got as far as I could with the principle of non-sufficient reason, but it turns out in some cases to give answers quite contrary to general belief. E.G. as Broad pointed out, it will never give a reasonable probability to a general law of the form 'all crows are black'. He tried to get a workable alternative but got no answer very convincing either to him or to me ; I discussed the point very shortly in Scientific Inference 191-197, but I think the attempt in my Camb.Phil. Soc. paper on sampling is on the right lines. The case of quantitative laws is even more to the point. E.g., in the case of uniform acceleration discussed in Scientific Inference 37-41, we could choose an infinite number of laws that would fit the data perfectly, but nevertheless do choose one, on a priori grounds, that only fits them approximately. If likelihood was the only thing that mattered we would choose any of the exact solutions in preference to the one we actually do choose.

The principle of non-suffiement reason is intended to serve simply as an expression of lack of prejudice ; in a sampling problem we want to give all constitutions of the whole class an equal chance of acquiring a high probability by experiment. But in these cases of general laws there seems to be prejudice ; I cannot help it, but there is a general belief in the possibility of establishing quantitative laws by experience, and I am not prepared to say that the general belief is wrong. I think I have stated a postulate that expresses it sufficiently clearly for practical purposes, but the postulate cannot be proved experimentally. But since such a postulate must from its nature be believed independently of experience that is a recommendation. What I want is, since an <u>a priori</u> postulate is needed anyhow, to choose it in such a way that the maximum number of alternatives are left for experience to select from. Eddington, e.g., chooses it too drastically by saying that the law of gravitation <u>must</u> be generally covariant and that space <u>must</u> have a finite curvature, deliberately excluding ~~xkm~~ $\phi$ 0 from the admissible values of a certain coefficient.

Your paper on Inverse Probability in the C.P.S.Proc. involves a subtle point. Suppose we have a huge population and from it make up classes each of 10000 members, containing respectively 0,1,2 .... with the property $\phi$. We choose one of these at random and select a sample of 100, of which 30 have the property $\phi$. You say, and I agree, that this establishes a high probability that the class chosen contains about 3000 with the property. If on the other hand we chose 10000 at random, sampled them as before with the same result, we should still estimate by inverse probability, using n.s.r., that the class contained about 3000 $\phi$'s. You ~~xxm~~ think it absurd that the two should give the same result. Actually ~~xx~~ they do when you only sample one class ; but if you take two classes there is a difference. In the first case the fact that one class has been estimated to contain about 3000 $\phi$'s slightly reduces the probability that the second will contain a number in this range, because one of them is excluded from the possible compositions of the second. But in the second case the first sample of 100 is effectively a sample of the whole super-population, and establishes that the ratio in this is probably about 3 to 7. Hence there is an <u>increased</u> prior probability that the second 10000 sampled will be in about this ratio. If we again get 30, we shall feel increased confidence that the number is about 3000 ( and reduce the probable error) : but in the first case what change there is is in the other direction.

I don't understand your remark about a verbal point in your criticism of my remarks on Keynes. Mises said something similar in reviewing my book. My immediate reaction is that I cannot take it as obvious that people's linguistic habits are meaningless ; but I think there is something else and can't see what it is.

I keep on getting other jobs pushed on to me and have not yet managed to read your paper carefully, though I have re-read a good deal of previous work (damn the spacer on this machine ; it keeps jumping o or 2 when I want it to do 1). In your discussion of my 1/3 business you seemed to assume that the third observation was equally distributed in some way ; this is contrary to the postulate that the probability follows the normal law. I don't think there is much to add to Bartlett's discussion. My own got rather muddled through getting statements in the wrong order. The point is that if we know anything about <u>h</u> to begin with, and the first two observations are separated by a big multiple of 1/h, there will be an extra probability that the third will lie between them ; if they are separated by a small multiple, it will be probable that the third will lie outside them by a distance of order 1/h. When the first two are fixed, the probability that the third will be between them is 1/3 only if we have no previous knowledge about h.

I will think about your problem of the trees.

Ithink that as a result of this you may feel inclined to alter your paper somewhat ; if I answer it as it stands I think I shall miss the important points and say too much about others. Should I send it back for any modification you think desirable and then I can have another try ?

Yours sincerely,

Harold Jeffreys.