St John's College
Cambridge.

March 21. 1934

Dear Fisher,

I have got this more or less done at last ; sorry to be so long. I have not typed it as there may be further revision.

I think it may help readers if we could put things the other way round. I have arranged my paper so that it will be an explanation of my own point of view ; how far the opinions I have dealt with are yours I don't know, but as they are some people's I think they should be discussed. I ~~dzmkx~~ think your paper is not relevant to my real arguments and that as it stands it will tend to obscure things; so I suggest that mine might come first and you might add comments at the end. That of course is as you like. I could put a note at the beginning to say that this is the result of discussion between us.

I have thought a bit more about your question about the wild plants, but am hazy about what you want. If it is a sampling problem, total number of individuals n, with $r_1, r_2 \ldots$ of the various species, I think that $P(r_1, r_2 \ldots | n)$ should be constant for all values subject to the sum being n. This seems the obvious generalization of Laplace. W.E.Johnson had a paper in Mind about a year ago (posthumous, edited by Braithwaite) in which he did something of the sort ; the remarkable thing about it was that he did the whole thing on the posterior probability, but he had an undetermined constant left over. That's the trouble about philosophers ; they talk sense up to a point (some of them I mean) but run away when they're on the verge of saying something useful. On the other hand you may be thinking of a Mendelian problem where there is one species and the question is about the incidence of one factor. Then I think my C.P.S. paper is relevant. But I may have missed the point altogether because I couldn't see where your ratios $1, r, r^2 \ldots$ could have come from.

Frank Ramsey seems to have had the gift of thinking clearly about these things more than anybody I know, though I have hopes of Broad. If somebody of that type would get on to it Ithink we could make some progress. I'm not desperately interested myself, being willing to take approximations and even orders of magnitude when it is too much trouble to get anything better ; but if,e.g. the simplicity postulate could be stated more precisely than I have done it would serve as a better guide than we have yet as to when we ought to introduce a complication. The sort of thing that bothers me is this. In seismology we get times of transmission to various distances, and fit a polynomial of degree 3, say, to them. The significance of the last term really involves the prior probability that such a term will be present. The usual thing is to keep it if it is some arbitrary multiple of its standard error, but Ithink it ought to be possible to frame a rule with <u>some</u> sort of argument behind it.

By the way you may be able to help me about the following. I was discussing the variation of drum-rate of seismographs lately, the data being the distances turned by the drums in consecutive minutes. I usually took about 16 observations. The correlations in, as near as I can remember, 3 instances out of 12 were negative and one was over 0·25. Theoretically they should all be positive. According to Yule the standard error of r is $(1-r^2)/\sqrt{n}$ or about 0·25. Thus 2/3 of the observed values agreed within half the standard error. Is there a catch somewhere or was I lucky ?

I want some time to tackle the problem of the distribution of gravity over the earth. The trouble is that the stations are very irregularly distributed and there is no question of being able to interpolate directly and work out coefficients of spherical harmonics by integration. The nearest I can see to a practical method is to assume a series of harmonics up to order 4 and try to fit it to the observations. But then the harmonics of order 5 may have the same sign over most of the United States, and it would be ridiculous to treat each station there with th same weight as a single station in the Antarctic Continent. Some kind of grouping seems indicated ; the question is what would be the best area to include in a group ? Somebody may have had a similar problem in trying to fit a Fourier series to a gapped series of observations.

Yours sincerely,

Harold Jeffreys