November 8th, 1935

My dear Irwin,

I am glad you have give me an opportunity of re-considering the arguments in the neighbourhood of page 721 of my paper on Statistical Estimation. The main point, which I think only becomes clear when the paper is read as a whole, is that we are concerned with the difference between two systems of equal statistical surfaces or regions, and usually not with the difference between different regions of the same system. I am returning your letter for reference, together with a few sheets which I hope clear up the main difficulty, though I don't discuss the analytic handling of the variance of the second degree deviations between corresponding surfaces of the different systems. I will go into this with pleasure if you would like me to.

Yours sincerely,

1. The origin of the loss of information incurred by using an estimate from a sample in conjunction with the sampling distribution of all such estimates from samples of a given size, has been traced to the fact that the equistatistical surfaces on which

$$\frac{\partial L}{\partial \theta} = 0,$$

for different values of θ do not coincide with the system of surfaces on which

$$\frac{\partial L}{\partial \theta} = c,$$

for any one value of θ. The systems clearly have one surface in common, and the neighbouring surfaces of the two systems will generally be inclined to each other at only small angles; e.g., if

$$\hat{\theta} - \theta$$

is small, the surface of the first system

$$\frac{\partial L}{\partial \hat{\theta}} = 0$$

may be chosen, with a certain harmless arbitrariness, to correspond with one of the second system, such as that which meets it on the locus of expectation

$$\frac{\partial L}{\partial \theta} = S\left(\frac{\hat{m}}{m}\frac{\partial m}{\partial \theta}\right).$$

Then my assertion is that the angle between these corresponding surfaces of the two systems will, in the neighbourhood where $\hat\theta$ is near to θ be proportional to the difference $\hat\theta - \theta$

The equations of the two surfaces are in fact

$$S\left(\frac{x}{\hat m}\frac{\partial \hat m}{\partial \theta}\right) = 0, \text{ and } S\left(\frac{x}{m}\frac{\partial m}{\partial \theta}\right) = S\left(\frac{\hat m}{m}\frac{\partial m}{\partial \theta}\right)$$

and the coefficient of x in the first may, if the differential coefficients of log m are finite, be expanded as

$$\frac{1}{m}\frac{\partial m}{\partial \theta} + (\hat\theta - \theta)\frac{d^2}{d\theta^2}\log m + \frac{(\hat\theta-\theta)^2}{2}\frac{d^3}{dm^3}\log m + \cdots$$

using the convention that ~~the squa~~ the square of the cosine of the angle between surfaces

$$S(px) = C$$

and $$S(p'x) = C',$$

is

$$\cos^2\alpha = \frac{S^2(mpp')}{S(mp^2)S(mp'^2)}$$

one can obtain

$$s^2_\alpha = \frac{\hat{\theta}-\theta^2}{I} \; S\left\{\frac{1}{m}\left(m'' - \frac{m'^2}{m}\right)^2\right\}$$

2. It being agreed that we have to evaluate the variance of

$$\frac{\partial^2 L}{\partial \theta^2}$$

over regions for which

$$\frac{\partial L}{\partial \hat{\theta}} = 0$$

the question is whether it is sufficient to evaluate it within the regions for which

$$\frac{\partial L}{\partial \theta} = C$$

Might one not argue as follows? - The variance within any one of the regions

$$\frac{\partial L}{\partial \hat{\theta}} = 0$$

will be the same as that evaluated, together with a corrective term proportional to $\hat{\theta} - \theta$ and other terms of higher order. Averaging all these according to their frequency of occurence in the limit for large samples, we obtain a corrective term proportional to the variance of

$$\hat{\theta} - \theta,$$

or of order 1/n compared with the value obtained.

I think this also covers the logical, though not the analytical difficulty on your p.3.

With respect to p.4 one might, perhaps, put the position in this way:- By basing our notion of precision only on a single estimate and the size of the sample, we are not neglecting such information about precision as our estimate provides, but we are neglecting other information provided by the sample respecting the precision of an estimate of given magnitude derived from it. Common formulae for standard errors, for example, are of the form

$$\sigma = b(\theta, n)$$

Consequently, variations in precision associated with variations in θ are taken into account throuth the magnitude of our estimate, but variations in precision for samples of the same size and giving the same estimate are neglected.