15 February 1932.

Prof. J.B.S. Haldane, M.A.,
John Innes Horticultural Inst.,
Mostyn Rd.,
Merton Park,
London, S.W. 19.

Dear Haldane:

There are really two parts to the question of efficiency in the problem you send, efficiency in the treatment of a group of families of the same size, and efficiency in the combination of information derived from families of different sizes. You do not mention it specifically, but I imagine that Bernstein sums $s(s-1)$ for all families, and if so this would determine the second part of the question.

On the first point Bernstein's method is, as you suspected, not efficient, but comes remarkably near to being so. Families of one clearly give no information and families of two and three are each divisible into onlt two classes, according as all or one less than all the children fall into one or other of the cross-over versus non cross-over groups. With material of this

class any consistent statistic must be efficient, since a cross-over value can be found to fit exactly any particular ratio between the two classes that may be observed. (in fact the data are reproducible from the estimate) In families of four and five where there are three classes we get the first real case involving combination of observations in which no estimate will give an expectation exactly reproducing the data. I have worked out the efficiency of Bernstein's method for the values of p in the following table, using that symbol for the recombination fraction.

Efficiency for families of the same size.

p	Size of family 4	5	6
0	100.0	100.0	100.0
0.1	95.677	95.921	95.676
0.2	95.365	94.307	93.129
0.3	97.618	95.938	94.461
0.4	99.691	99.231	98.704
0.5	100.0	100.0	100.0

The value for p = 0.1 for families of four looks too low and may contain an error. I went as far as families of six in case on increasing from three classes to four there was an appreciable drop in the efficiency, but evidently there is not, and I imagine that if efficiencies below 90 per cent. occur they will only affect a very small fraction of the data, contributed by a few of the largest families.

I have not hunted up if Bernstein gives his data, but if he does it would not be a long job to recalculate the estimates by maximum likelihood. I should be inclined to handle the whole thing arithmetically, for though the algebra is not too heavy up to families of five, it does become heavy thereafter, and the arithmetic is particularly pleasant in exhibiting the amount of information actually contributed by the families of different sizes and weighting them automatically. Formally, the algebra involves solving equations of degrees somewhat higher than the size of family and this is anyhow best done by trial. If Bernstein does not give the data I would, if you liked, exhibit the arithmetical process on fictitious material, so that you could see how it runs out.

I have not seen Jeffrey's book and I suppose that you are right that I ought to, but I do not suppose that he comes very close to the real problems.

Yours sincerely,