27th. February 1947.

Dear Thornton,

I have looked through the paper by Jones and Mollison and the appendix by Quenouille, and am immensely interested to see how competently the new dried agar method deals with the difficulties of direct enumeration of soil flora.

The practical question which arises from the occurrence of bacterial cells in clutches or clumps lies, of course, in the estimation of the precision of a given count. I do not think that Quenouille has said what to do about this, and perhaps the matter has not been brought to his attention; but it is clear that the samples should not be treated as random samples from a negative binomial, since in the enumeration both the clumps and the numbers in each clump are enumerated, and it would be foolish to throw away this additional information as to the precision of the count.

In the case of a Poisson distribution, as for single dispersed particles, the variance of the count N may, as we have all known [illegible], be taken to be N itself. For a count based on a Poisson distribution of counts, the cells having a logarithmic distribution of number per clump, the sampling variance will, I think, be $\frac{N}{1-x}$, i.e. about four times as great for a given number of cells counted as the ordinary formula, to judge by the values of x fitted to your series.

I have no doubt Quenouille can verify this if he looks at the problem from this point of view, recognising that all the information in the count is to be used in the estimation of error.

This formula depends on the logarithmic series; and, although I think the logarithmic series is sufficiently closely realised for this purpose, yet there are in the figures given in the paper discrepancies which I think should not escape attention.

In the four tables given for the shaking test, in which observed and expected frequencies are compared, it is striking that the numbers observed are greater than expected for clumps of two, less for clumps of three and greater for clumps of four, to an extent which can scarcely be due to chance. This may, I suppose, well be due to reproduction by binary fission producing pairs of cells obstinately adherent in spite of shaking. Anyway these four tables give

Clump size		
1	57.3	too few
2	80.0	too many
3	44.0	too few
4	28.2	too many
5	5.8	too few
6	0.8	too many
7	1.9	too few

There is a similar ~~distribution~~ discrepancy of lower magnitude in the series given for samples A, B, C, C_1, C_2. Here there seems to be a

~~discrepancy~~ error in the total for sample B, which I should guess might be due to the expected number for colonies of three bacteria being put down as 413 instead of 493.

Jones also asks about ~~a~~ this theoretical derivation of the logarithmic series which I think I understand apart from his introduction under (c) of the "simplest possible damping effect (due to senescence, autolysis, competition etc.) which is a power law". I feel that this is rather less than cogent as a basis for introducing the term x^n, where n is the number of cells in the colony. I do not know whether the damping effect is conceived of as acting to change the length of time during which a growing colony retains the same number of individuals or whether it is supposed to act in some other way. However, it is certain that the logarithmic series does arise in an immense variety of circumstances, and presumably, therefore, many different mathematical situations must lead to it.

Yours sincerely,