July 7, 1938

Dear Koshal,

I will answer first the genetical part of your letter. If \( s \) parent lines had been used with the complete set of \( \frac{1}{s}(s - 1) \) first cross progenies, one could pick out \( s - 1 \) comparisons among the \( \frac{1}{s}(s + 1) \) sets of samples available, using the form

\[
\begin{align*}
2 & \text{AA} + \text{AB} + \text{AC} + \ldots \ldots \ldots \ldots \ \text{AZ} \\
2 & \text{BB} + \text{AB} + \text{BC} + \ldots \ldots \ldots \ldots \ \text{BZ} \\
& \text{s items}
\end{align*}
\]

These would compare the effects of the whole sets of genes A, B, etc, characteristic of the \( s \) parent strains. The comparison enables one to say which varieties give generally the best results on crossing.

In addition the material gives \( \frac{1}{s}(s - 1) \) comparisons of the form

\[
\text{AA} + \text{BB} + 2\text{AB}
\]

i.e., the double value of each cross may be compared with the sum of the performance of the two parent lines. The effect known as heterosis is that, in some species and in some
characters these comparisons would be predominantly negative, consequently their total carries a single comparison for heterosis, or, as it may be called, for dominance bias.

There remain the $\frac{3}{4}(s-2)(s+1)$ comparisons representing the variation among the last line lot of $\frac{3}{4}s(s-1)$. In Calcutta I think I spoke of these as due to epistasy, but this is a wide use of the word, and it is difficult to name the effect, if it exists, in genetic terms. Since in each comparison direct additive effects of the genes are eliminated, it clearly can only depend on the way different genes interact, and this is generally spoken of as epistasy.

I liked your analysis of the three cotton lines, showing in that case that the genetic comparison alone explained the observations, neither heterosis nor epistasy having any appreciable effect. I think this may often turn out to be the case, but the plant breeder will find it useful when departures from such a simple rule are indicated.

I do not think it is surprising that the discrepancies from the conditions of binomial $X^2$ should increase as the approximation to maximal likelihood improves, seeing that your figures show increasing values of $X^2$ from your first to your third approximation. Evidently your trial values
lie between maximal likelihood and minimal $X^2$, as I think also the true moment solution would do.

I think I should deduct all the values of log-likelihood from its absolute maximum minus 944.9986807, i.e., sum of $a \log \frac{a}{n}$ where $a$ is the frequency observed in any class. One then can make such comparisons as

<table>
<thead>
<tr>
<th>Deficiency of Log likelihood</th>
</tr>
</thead>
<tbody>
<tr>
<td>According to Pearson's figures</td>
</tr>
<tr>
<td>Maximal likelihood</td>
</tr>
<tr>
<td>Pearson's method</td>
</tr>
<tr>
<td>Difference</td>
</tr>
</tbody>
</table>

parallel with the similar comparisons for $X^2$

| According to Pearson's figures | Corrected values |
| Maximal likelihood           | 9.1453          | 6.8748          |
| Pearson's method             | 7.6820          | 8.7980          |
| Difference                   | 1.4633          | -1.9232         |

In the above log I have used the values on your sheet 1, namely 7.6820, for Pearson's solution, but I find in my paper the value 7.6922 for what Pearson's frequency, if correct, would give. Perhaps you will look into this discrepancy. In any case you will see that, by subtracting the log likelihood from the absolute maximum, one obtains a comparison analogous
to that supplied by $X^2$, and giving in fact very similar results. This is by way of commentary, upon Pearson's foolish complaint that the value of log. likelihood were less sensitive than those of $X^2$.

I am returning your summary pages, and shall be glad to see your manuscript when you have drafted it.

Yours sincerely,