November 17, 1941

Dear Koshel,

Thanks for your letter. On the point on which you ask my advice, it may be worth noting that the standard form given by Pearson for the Type IV curve, namely:

\[
\frac{1}{F} \tan^{-1} x (1+x^2)^{-\frac{1}{2}} \left( \frac{1}{2} + 2 \right) \tan^{-1} x
\]

is really an analytic transformation of Type I, which you have already dealt with, since

\[
\frac{1}{2} \log \frac{x-1}{x+1} = -\frac{1}{2} \pi + i \tan^{-1} x
\]

so that the frequency formula may be written

\[
(x+1)^{-\frac{1}{2}} (r+2)^{-\frac{1}{2}} (x-1)^{-\frac{1}{2}} (r+2)^{\frac{1}{2}} \tan^{-1} \frac{x}{\beta} \frac{dx}{F}
\]

The two constants of the curve other than \( \beta \) and \( r \) are of course those concerned with location and scale, and are introduced by a transformation

\[
x = \frac{y - \mu}{\beta}
\]

The methods you are illustrating enable one to approach the maximal likelihood fit, for in material in which the expectations \( m \) can be calculated for any set of trial values of the parameters; but the calculations of expectations for Type IV are at least as laborious.
as those for Type I, and in view of this analytic equivalence there is very little to be said for working an example of this type.

An arithmetical approach to fitting problems in which the data have such peculiarities as

1) frequencies for $x$ less than $a$ unknown and unobservable;

2) frequencies for $x$ greater than $b$ pooled in a single frequency, possibly a large fraction of the total frequency observed, and which is therefore applicable in many cases where the method of moments could not even be attempted, is as follows:

For a trial value $\theta$ of an unknown parameter, expectation $m_1$ are calculated for the observable frequencies, and expectations $m_2$ for the second neighbouring value $\theta_2$. For each observable class we may then calculate

$$\frac{m_2 - m_1}{\theta_2 - \theta_1} \cdot \frac{2}{m_1 + m_2} = w$$

Then $I = Sw \left( \frac{m_2 - m_1}{\theta_2 - \theta_1} \right)$

gives the approximate amount of information available about
when all other parameters are known, at the values assumed, and

$$\theta_1 + \frac{Sw(a - m_1)}{Sw(m_2 - m_1)} \cdot (\theta - \theta_1)$$

is an improved value of $\theta$ based on the two first trials. If $\theta$ is a second parameter, the same process may be carried out at a second trial value for $\phi$, and an improved value of $\phi$ found from the two distributions, using for each $\phi$ nearly the best value at for $\theta$ as found from the two trial values for $\theta$. When one sufficiently good set of values has been found in this kind of way (p.t.o.}
of course in a position to make a really accurate fit, using the method you developed in your first paper on fitting.

With best wishes,

Yours sincerely,