Dear Prof Fisher.

I should be very grateful for your comments on the enclosed. I would ~~have discussed~~ like to discuss it with you sometime – but I thought perhaps it would be easier if I wrote out my argument more or less formally first.

yours sincerely

J H Gaddum

THE FITTING OF CURVES TO TOXICITY DATA

By J.H. Gaddum

(University College, London, W.C.1.)

THE FITTING OF CURVES TO TOXICITY DATA

By J.H. Gaddum

(University College, London, W.C.1.)

An important group of pharmacological tests depends
on fitting curves to data connecting the dose of a drug
with the percentage mortality, (Trevan, 1927).   When the
doses are measured on a logarithmic scale these data can
usually be fitted by the curve representing the integral
of the normal frequency distribution (Gaddum, 1933).
The most convenient way to do this is to convert the observed
mortalities into suitable units ("normal equivalent deviations"
or "probits") and to fit a straight regression line.   This
may be done either graphically or algebraically.

In 1935 Bliss published two papers covering some of the
ground already covered by Gaddum and ~~purporting to~~ introducing
improved methods.   The object of the present note is to show
that one of the modifications proposed by Bliss not only
complicates the computation, but also diminishes the accuracy
of the result.

In data of this type the weight of an observation is
maximal when the mortality is 50 per cent and becomes zero

when the mortality is either 0 or 100 percent.   It might
appear superficially that the most accurate results would
be obtained by first fitting a regression line roughly to
the observed results and then weighting each observation
according to the estimate of the true result obtained
from this roughly fitted line.   This method was considered
by Gaddum, but rejected on the grounds that a more accurate
estimate could be obtained by calculating the weight of each
observation from the result actually observed.   Bliss rejects
this method without discussion and reverts to the method
rejected by Gaddum.   Gaddum's original argument was
considerably compressed owing to editorial obstruction.   This
argument, which is amplified below, is based on the method of
maximum likelihood.
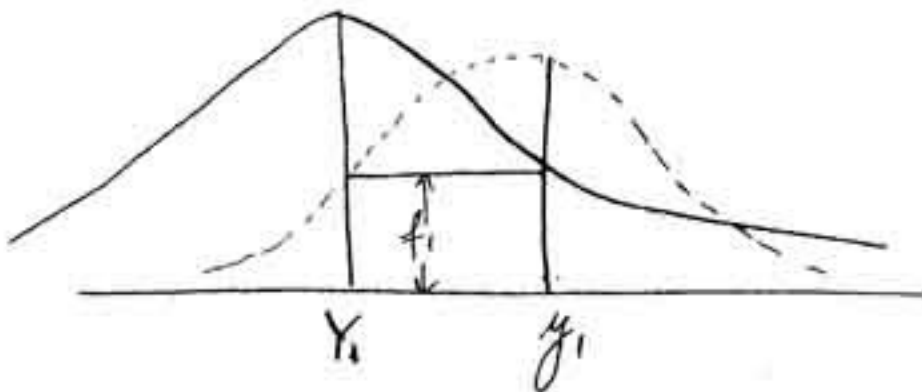
In the figure

Y_1 is the true value of a variable.

y_1 is the observed value.

The continuous line is the probability distribution of y_1 about Y_1.

The broken line is the "curve of likelihood" showing the likelihood of different values of Y_1 when y_1 is given.

f_1 represents the likelihood.

Neither of these curves is necessarily normal.

The method of maximum likelihood involves finding the maximum value of $\sum (\log f_1)$ and this is equivalent to finding

the minimum value of $\sum (-\log f_i)$. Fisher (1921) has given reasons for believing that this gives the best possible estimate. If this is so the value of other methods of calculation can be judged by testing whether or not they are equivalent to this method.

The method of least squares consists in finding the minimum value of $\sum \frac{(y_i - Y_i)^2}{\sigma_{Y_i}^2}$ where $\frac{1}{\sigma_{Y_i}^2}$ is the weight of the observation.

This will be equivalent to the method of maximum likelihood if $\sum (-\log f_i)$ is proportional to $\sum \frac{(y_i - Y_i)^2}{\sigma_{Y_i}^2}$

i.e. if $\log f_i = K - \frac{(y_i - Y_i)^2}{\sigma_{Y_i}^2}$

or $f_i = K e^{-\frac{(y_i - Y_i)^2}{\sigma_{Y_i}^2}}$

i.e. if the probability distribution is a normal curve with standard deviation σ_{Y_i}

If the probability distribution is not normal the two methods will not, in general, give the same result.

The method of least squares is applied by Bliss in this form to data where y represents the probit. This would give the same solution as the method of maximum likelihood, if the probability distribution of probits were normal. When the number of animals in each group is infinite this is true, but

When the groups are of the size used in actual experiments it
is not.   For example, if there are 20 animals in each group
and the true mortality is 20 per cent, the probability
distribution of probits (Gaddum, 1933, page 41, curve D) differs
from a normal curve in two important respects.

(1) It is discontinuous.

(2) It is grossly asymmetrical and doesn't reach the
base line even at $y = -\infty$

The new method recommended by Gaddum is based on the
surprising discovery that the curve of likelihood (Curve B
in the same figure) is an extraordinarily close approximation
to a normal curve.   The formula for the standard deviation
of this curve is the same as the formula which gives an
approximation to the standard deviation of the probability
distribution of probits.   The important point is that the
probit value used to estimate this standard deviation is y_1
and not Y_1.

If the curve of likelihood is taken as a normal curve
with standard deviation σ_{y_1}

$$f_1 = \frac{1}{\sqrt{2\pi}} e^{-\frac{(y_1 - Y_1)^2}{\sigma_{y_1}^2}}$$

or

$$-\log f_1 = \frac{(y_1 - Y_1)^2}{\sigma_{y_1}^2} + a\ constant$$

The solution given by the method of maximum likelihood is that corresponding to the minimum value of $\sum(-\log f_i)$

This is equal to that corresponding to the minimum value of $\dfrac{(y_i - Y_i)^2}{\sigma_{y_i}^2}$

The new method depends on finding this solution.

The thing which makes this more accurate than the method used by Bliss is that the weight of an observation is $\dfrac{1}{\sigma_{y_i}^2}$ and not $\dfrac{1}{\sigma_{Y_i}^2}$ Bliss attempts to estimate σ_{Y_i} His result would be less accurate than Gaddum's even if he knew σ_{Y_i} exactly.

REFERENCES

Bliss, C.I. (1935) Annals of Applied Biology., 22, 134 & 307.

Fisher, R.A. (1921) Philosoph. Trans. Roy. Soc. A., 222, 309.

Gaddum, J.H. (1933) Med. Res. Coun. Sp. Report., No. 183.

Trevan, J.W. (1927) Proc. Roy. Soc. B., 101, 483.