Maximum Likelihood fit to Points originating from different Poisson Distributions

by D.J. Gorman and J.W. Müller

Bureau International des Poids et Mesures

1. Introduction

In the determination of absolute alpha-particle energies in a magnetic spectrograph the energy is obtained by extrapolating the high energy edge of the alpha line to zero intensity (see, for example, ref. [1]). The method consists of fitting to the experimental points a theoretical intensity distribution I_x proportional to $(x - a)^{3/2}$ where x is the position on the photographic plate and a is the intercept. The procedure used until now [2] has been to take I_x , the measured number of alpha tracks per unit interval, subtract from it the average background I_o and transform the resultant quantity $I_x - I_o$ to the power 2/3. Then, using appropriate weights, a straight line of the form $x = a + b(I_x - I_o)^{2/3}$ is fitted by the method of least-squares. This procedure has the advantage that it leads to simple analytic expressions [2] for the intercept a and its variance σ_a^2.

The total number of measured tracks I_x follows a Poisson distribution and hence its variance is given by $\sigma_x^2 = I_x$. When setting error limits on this number, one commonly writes the best value as $I_x \pm \sigma_x$. However, this does not reflect the asymmetry of the Poisson distribution for which one would expect asymmetric error limits. For large values of I_x , this asymmetry becomes quite small and can then be neglected, but in the region where we are actually performing the fit (i.e. at the high energy edge of the α-line), I_x approaches zero.

It would therefore seem that a more accurate extrapolation of the α-line to zero intensity would be obtained by considering the Poissonian nature of the process and the fact that the errors about the measured numbers of alpha tracks will be asymmetric. The technique of least-squares is unable to treat this case of asymmetric errors and so a procedure has been adopted here based on the principle of maximum likelihood.

2. Determination of Asymmetric Errors

A treatment of the problem of determining confidence limits on rare events distributed according to the law of Poisson is given by van der Waerden [3] and we only quote here his results.

If k is the observed number of events of a process following a Poisson distribution law, then the confidence limits on the value k are given by

$$\lambda_{\pm} = k + \frac{1}{2} g^2 \pm g(k + \frac{1}{4} g^2)^{1/2} \quad , \tag{1}$$

where λ_+ and λ_- are the upper and lower confidence limits, respectively. The factor g determines the degree of confidence and is taken from a normal distribution. Thus, for example, g = 1 corresponds to confidence limits of about 68%. If the distribution were normal this would agree with the usual standard deviation.

For small values of k the limits of (1) differ considerably from those obtained by simply taking $\sigma = \sqrt{k}$. A few values are given in the table below for comparison. The value of g is taken as unity.

Table 1

Error limits for small total counts (g = 1)

Number of counts (k)	Lower limit from (1)	Upper limit from (1)	$\sigma = \sqrt{k}$
0	0.00	1.00	0.00
1	0.62	1.62	1.00
2	1.00	2.00	1.41
3	1.30	2.30	1.73
4	1.56	2.56	2.00
5	1.80	2.80	2.24
6	2.00	3.00	2.45
7	2.20	3.20	2.65
8	2.37	3.37	2.83
9	2.54	3.54	3.00
10	2.70	3.70	3.16
20	4.00	5.00	4.47
30	5.00	6.00	5.48
40	5.84	6.84	6.32

The question now arises as to what is the best fit of our theoretical line shape to points with asymmetric errors.

3. Principle of Maximum Likelihood

The method used in what follows is based on the principle of maximum likelihood. The measured points are designated by (x_i , y_i), where x_i is the position of the centre of the measured interval on the photographic plate, and y_i is the number of tracks counted (including background) in this interval. We assume that there is no error in the x_i , and that the error distribution in the y_i is Poissonian and thus given by

$$p(y_i \,|\lambda_i) = \frac{\lambda_i^{\,y_i}}{y_i!} \, e^{-\lambda_i} \quad . \tag{2}$$

λ_i is the (unknown) true value of which y_i is a measure.

If λ_i were known, we could calculate, using (2), the probability of observing a certain number of tracks y_i . However, we have only the measured quantity y_i , and wish to use it to estimate λ_i by means of the inverse probability $p(\lambda_i \,|y_i)$, for the distribution of the unknown true values λ_i.

This can be done by applying the method of Bayes [4] which gives

$$p(\lambda_i \,|\, y_i) = \frac{p(\lambda_i) \, p(y_i \,|\, \lambda_i)}{\int_0^\infty p(\lambda_i) \, p(y_i \,|\lambda_i) \, d\lambda_i} \quad . \tag{3}$$

Assuming constant a priori probabilities for the λ_i , the integral in the denominator of equation (3) is

$$\int_0^\infty p(y_i \,|\lambda_i) \, d\lambda_i = \frac{1}{y_i!} \int_0^\infty \lambda_i^{\,y_i} \, e^{-\lambda_i} \, d\lambda_i = \frac{1}{y_i!} \, \Gamma \, (y_i + 1) = 1 .$$

Therefore

$$p(\lambda_i |y_i) = \frac{\lambda_i^{\,y_i}}{y_i!} \, e^{-\lambda_i} = p(y_i \,|\lambda_i) \quad . \tag{4}$$

The density (4) has a maximum at $\lambda_i = y_i$. The measured y_i is therefore the most probable value of λ_i and for each y_i there belongs a distribution in λ_i of the form of equation (4).

If we now attempt to fit a family of curves to N points, each having y_i characterized by equation (4), the best fit, by the principle of maximum likelihood, will be obtained when the product P of the probabilities for each point λ_i is a maximum, i.e. when

$$P = \prod_{i=1}^{N} p(\lambda_i | y_i) = \text{a maximum.}$$

This gives

$$P = \prod_{i=1}^{N} \frac{\lambda_i^{y_i}}{y_i!} e^{-\lambda_i} = e^{-\sum_{i=1}^{N} \lambda_i} \prod_{i=1}^{N} \frac{\lambda_i^{y_i}}{y_i!} = \text{a maximum.} \tag{5}$$

Taking the logarithm this leads to

$$\ell n\, P = -\sum_{i=1}^{N} \lambda_i + \sum_{i=1}^{N} y_i\, \ell n\, \lambda_i - \sum_{i=1}^{N} \ell n(y_i!) = \text{a maximum .} \tag{6}$$

To proceed further we must specify the family of curves which we wish to fit to the data.

4. The form of the alpha line

We write the theoretical line shape (near the high energy edge) for the alpha line in our magnetic spectrograph in the following form (see [2]).

$$\lambda_i = a(b - x_i)^{3/2} + y_o \, . \tag{7}$$

In this equation y_o is the average background and a and b are constants to be determined. The constant b corresponds to the point on the plate at which the theoretical value λ_i becomes equal to background, and it is this quantity which is used to calculate the alpha energy.

The form of equation (7) is chosen in order to avoid subtracting the background from the measured points and taking the 2/3 power of the resulting counts, both of which would distort the Poissonian distribution of the y_i. The fit is performed directly to the measured data.

Substituting equation (7) into equation (6) gives

$$-a\sum_{i=1}^{N} (b - x_i)^{3/2} - N\, y_o + \sum_{i=1}^{N} y_i\, \ell n\left[a(b - x_i)^{3/2} + y_o\right] - \sum_{i=1}^{N} \ell n(y_i!) = \text{a maximum.} \tag{8}$$

Differentiating (8) with respect to a and b and setting the derivatives equal to zero, we obtain the two simultaneous equations

$$\sum_{i=1}^{N} (b - x_i)^{3/2} = \sum_{i=1}^{N} \frac{y_i (b - x_i)^{3/2}}{a(b - x_i)^{3/2} + y_o} \tag{9a}$$

and $$\sum_{i=1}^{N} (b - x_i)^{1/2} = \sum_{i=1}^{N} \frac{y_i (b - x_i)^{1/2}}{a(b - x_i)^{3/2} + y_o} \quad . \tag{9b}$$

Equation (9a) is obtained by differentiating (8) with respect to a and it is therefore the condition which, when fulfilled, gives the best value of a for b constant. Similarly, equation (9b) is obtained by differentiating (8) with respect to b, and its solution provides the best value of b for a constant. Unfortunately, there is no exact analytic solution, but equations (9) are readily solved numerically by means of a computer. One starts with trial values of a and b and by successive iterations a solution can be obtained to any desired degree of accuracy.

5. Error Estimation

Having obtained the best value of the intercept $b = b_o$ from (9), we would like to estimate its error due to statistical fluctuations in the data. Unfortunately, equations (9) prove just as resistant to solution for σ_b^2 as they do for b itself. However, the error in b may be estimated graphically by plotting distribution (5) as a function of b. The procedure is to fix a value of b, find the best value of a using equation (9a), calculate the set of λ_i using (7) and then calculate P from (5). Again, this is easily done with a computer. The resultant probability density will have a maximum at $b = b_o$. Confidence limits can be determined by taking the values of b which correspond to a probability equal to 0.607 of the probability at the maximum. For a normal distribution this would correspond to one standard deviation. These upper and lower error estimates will normally be different, reflecting the asymmetry of the initial Poisson distributions.

6. Example

This method has been applied to the analysis of a number of plates from the alpha-particle spectrograph. As an example Figure 1 shows the high energy edge of the alpha line obtained from a source of ^{240}Pu. The solid line is the calculated fit to the data. Figure 2 shows the calculated asymmetric distribution in the intercept with most probable value $(89.918\, ^{+0.024}_{-0.014})$ mm.

The corresponding energies from an analysis of four plates of ^{240}Pu are summarized in Table 2.

Table 2

Experimental values for the alpha energy of ^{240}Pu

Plate	Energy (keV) (least squares)	Energy (keV) (maximum likelihood)
242	5 168.35 $\pm$ 0.35	5 168.25 $^{+0.29}_{-0.14}$
243	5 168.24 $\pm$ 0.19	5 168.09 $^{+0.19}_{-0.11}$
244	5 168.35 $\pm$ 0.28	5 168.14 $^{+0.21}_{-0.16}$
246	5 168.43 $\pm$ 0.19	5 168.45 $^{+0.19}_{-0.18}$
weighted mean	5 168.34 $\pm$ 0.12	5 168.25 $^{+0.10}_{-0.10}$

Although on theoretical grounds, as explained above, one should expect this method to give more reliable results, extensive numerical calculations with data from our alpha-spectrograph show no significant difference between the results for the energies based on this and the usual least-squares method.

For the least-squares calculation the mean and standard deviations were calculated applying the usual formulae

$$\bar{x} = \frac{\sum x_i/\sigma_i^2}{\sum 1/\sigma_i^2}, \qquad \frac{1}{\sigma_{\bar{x}}^2} = \sum \frac{1}{\sigma_i^2} .$$

For the maximum likelihood calculation,an approximation was used as explained in the Appendix. The asymmetry, clearly visible in the errors of the individual measurements, tends to disappear in the mean, as might be expected on the basis of the Central Limit Theorem.

It is certainly reassuring that the two methods of calculation do not give widely disparate results and one can therefore, for most applications at least, apply the simpler least-squares method with confidence.

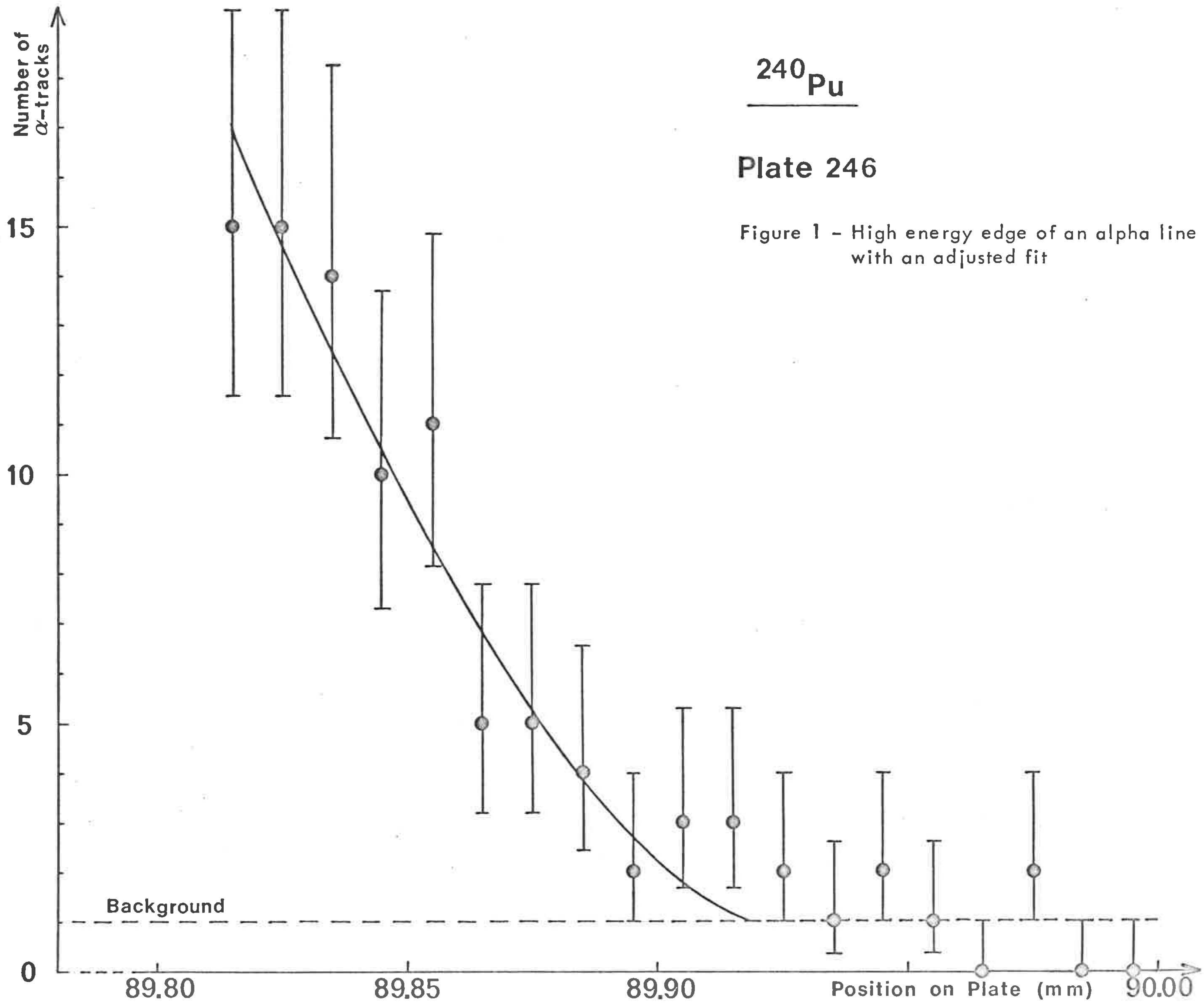

Figure 1 – High energy edge of an alpha line with an adjusted fit

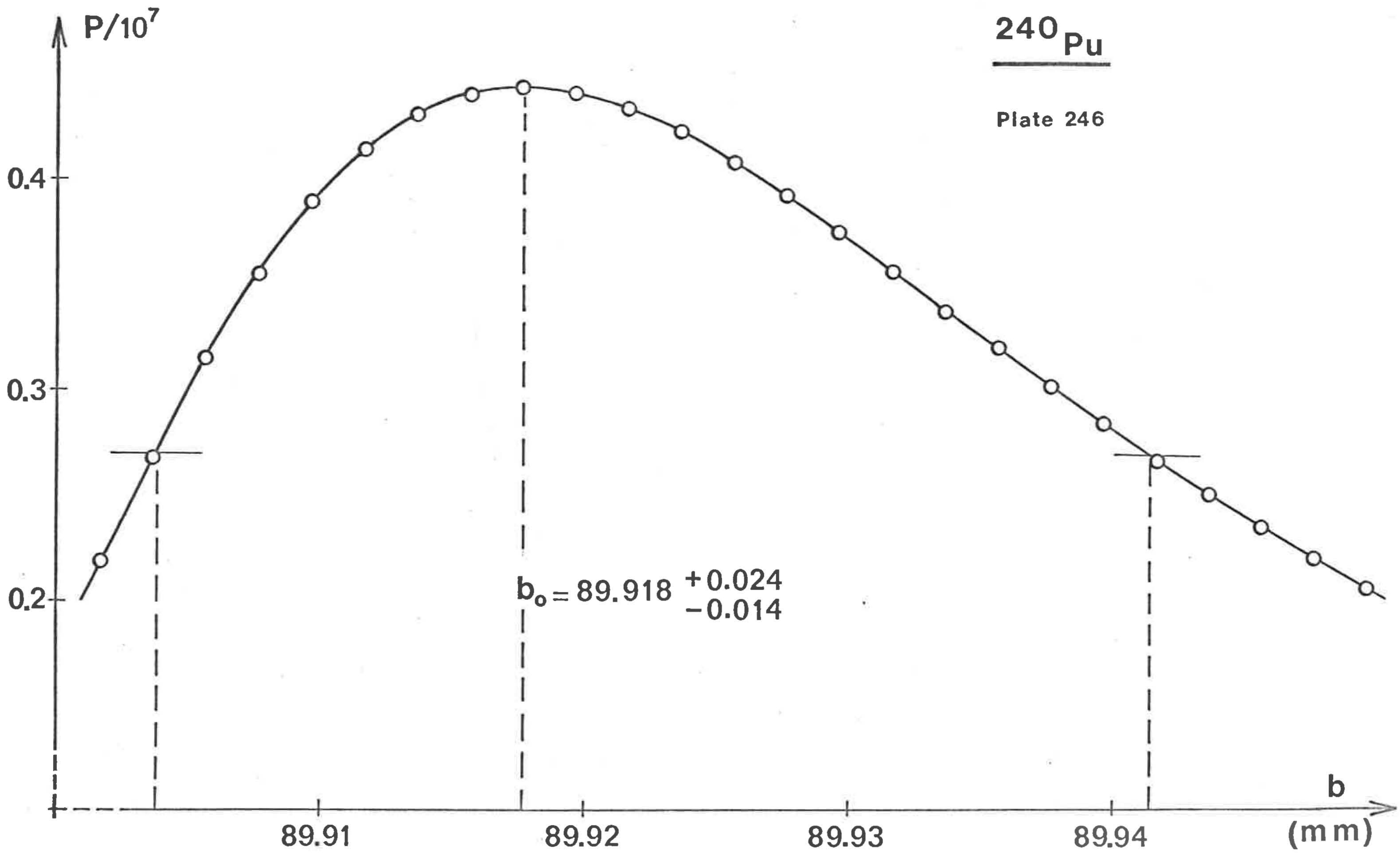

Figure 2 - Calculated probability distribution for the extrapolated end point b_o of an alpha line

Appendix

An approximate method for determining the "best" value from a set of measurements with asymmetric errors

Let us suppose that as a result of a series of N independent measurements on a quantity x we obtain the values x_1, x_2, ..., x_N with probability distributions $f_1(x)$, $f_2(x)$, ..., $f_N(x)$. The point x_i is that value of x for which $f_i(x)$ is a maximum. We wish to use the measured x_i to obtain a "best" value $\bar{x}$. This "best" value, in the sense of maximum likelihood, will be the value of x for which the function

$$F(x) = f_1(x) \cdot f_2(x) \cdot \ldots \cdot f_N(x) = \prod_{i=1}^{N} f_i(x) \tag{A1}$$

is a maximum.

In the case we are considering the distributions $f_i(x)$ have the form shown in Fig. 2 and must be evaluated numerically. In order to calculate F(x) exactly we would have to determine numerically all the densities $f_i(x)$ at a large number of points and then form their product. Although in principle this could be done, we have preferred to simplify the calculation by replacing the $f_i(x)$ by a "double gaussian" of the following normalized form.

$$g_i(x) = \begin{cases} C_i \cdot \exp\left[-\dfrac{(x - x_i)^2}{2\,\sigma_{i2}^2}\right], & \text{for } -\infty < x \leqslant x_i\,, \\[2ex] C_i \cdot \exp\left[-\dfrac{(x - x_i)^2}{2\,\sigma_{i1}^2}\right], & \text{for } x_i \leqslant x < \infty, \end{cases} \tag{A2}$$

with $C_i = \sqrt{2/\pi} \cdot (\sigma_{i1} + \sigma_{i2})^{-1}$.

Here x_i is the most probable value (b_o in Fig. 2) and σ_{i1} and σ_{i2} are the upper and lower confidence limits, respectively, as described in section 5 and shown in Fig. 2.

It should be noted that as a consequence of the asymmetry of equation (A2), the mean value of x

$$E_i(x) = \int_{-\infty}^{\infty} x \cdot g_i(x)\, dx = x_i + \sqrt{2/\pi} \cdot (\sigma_{i1} - \sigma_{i2}) \qquad \text{(A3)}$$

cannot be identical with x_i, unless $\sigma_{i1} = \sigma_{i2}$.

The "best" value $\bar{x}$ of x is therefore that for which

$$F(x) = \prod_{i=1}^{N} g_i(x) \quad \text{is a maximum.}$$

As before, the upper and lower confidence limits are the points x at which

$$F(x) \cong 0.607\, F(\bar{x}) \quad .$$

References

[1] A. Rytz: Helv. Phys. Acta 34, 240 (1961)

[2] B. Grennberg and A. Rytz: Metrologia 7, 65 (1971)

[3] B.L. van der Waerden: Mathematical Statistics (Springer-Verlag, Berlin, 1969)

[4] J.W. Müller: Traitement statistique des résultats de mesure, Rapport BIPM-108 (1969 ff).

(March 1972)