PUBLICAÇÕES

DETERMINATION OF THE UPPER LIMIT OF A PEAK AREA

O. Helene
Instituto de Física, Universidade de São Paulo

Março/1990

DETERMINATION OF THE UPPER LIMIT
OF A PEAK AREA*

O. Helene

Instituto de Física, Universidade de São Paulo,
CP 20516, 01498, São Paulo, S.P., Brazil

This paper reports the procedure to extract an upper limit of a peak area in a mutichannel spectrum. `.)`. procedure takes into account the finite shape of the peak and the uncertanties both i ʳ ᵉ ʾbackground and in the expected position of the peak.

I. INTRODUCTION

Many experiments in physics are running in order to indentify a peak in a multichannel spectrum. However, it is not rare that these experiments give negative results: the sought after peak is not clearly observed. This fact can occur when the peak is hidden in the statistical flutuation or in the uncertainty of the background or also hidden due to its own statistical flutuation. In some cases, the peak position is known with a non negligible standard deviation and we simple do not know exactly where search for it. In short, the peak can be lost in a statistical jungle. In these cases we can only determine the upper limit of the peak area with a chosen confidence level.

Extraction of the upper limit of a signal in the presence of a background, where both signal and background obey Poisson distributions, was already discussed in a previous paper (1). In this reference it was shown that when the background B is known with a negligible error. C events are observed in a bin of the multichannel spectrum, and a is the (unknown) intensity of the searched signal in the same bin, then the posterior probability density function of a is given by:

$$g(a) = N \frac{e^{-(a+B)}(a+B)^C}{C!} .$$

(1)

where N is a normalization constant such that:

$$\int_0^\infty g(a)da = 1 .$$

(2)

If $C > B$. that is. the peak is clearly seen. then $g(a)$ is peaked at a value $\hat{a} = C-B$ with a corresponding standard deviation equals to $\sqrt{C}$. However. there are many experiments

where $B \gtrsim C$. In this case, an upper limit A_α for a can be defined [1] by

$$\alpha = \int_{A_\alpha}^{\infty} g(a)\, da \quad ,$$

(3)

where $1-\alpha$ is the confidence level (C.L.). It can be shown [2,3] that eq. (3) is equivalent to

$$\alpha = \frac{e^{-(B+A_\alpha)} \sum_{n=0}^{C} (B+A_\alpha)^a/n!}{e^{-B} \sum_{n=0}^{C} B^a/n!} \quad .$$

(4)

There are three points that were not considered in ref. [1]: the shape of the peak in the multichannel spectrum, and the nonvanishing errors in the background (that were just considered in the limit $B > 1$ and $C > 1$) and in the expected position of the searched peak.

Recently Avignone [4] has also discussed this problem using a maximum likelihood procedure and including the shape of the peak. However, his results were restricted to a normal approximation of the Poisson distribution and statistical uncertainty of the background was not taken into account. Also ref. [4] does not consider the existence of a possible standard deviation in the position of the searched peak.

In this paper we will extend the results of ref. [1] in order to take into account the three points mentioned above. The results will include the case of small number of events.

II. ANALYSIS

a) Non vanishing background error

As stated in ref. [1], when B is not exactly known but obeys a probability density function $f(B')$, equation (1) must be rewritten as

$$g(a) = K \int_0^\infty \frac{e^{-(a+B')}(a+B')^C}{C!} f(B') \, dB' \quad , \tag{5}$$

where K is the normalization constant. It will be supposed in this paper that the knowledge of the background has a Gaussian shape with mean value B and standard deviation σ_B ,

$$f(B') \sim e^{-(B'-B)^2/2\sigma_B^2} \tag{6}$$

Using $f(B')$ above in eq. (5) and using eq. (5) in eq. (3) we can determine A_α. Figure 1 shows A_α as function of B for some C values, for $\sigma_B = 0$ and for $\sigma_B = \sqrt{B}$.

Figure (1) must be read as follows: if C events are observed in a multichannel region where the mean background is B with standard deviation σ_B , then A_α is the upper limit of the signal in the region with C.L. $1-\alpha$. As can be seen from figure (1) A_α is always positive also if C < B. If C > B that is, if the signal is actually observed, the upper limit A_α is compatible to the signal intensity C − B . We can also observe that for fixed B and C, when σ_B increases, A_α increases.

b) Finite resolution

If we are searching for a peak in a multichannel spectrum we must consider the finite resolution of the detector system. If the searched peak has a Gaussian shape centered at x_0 and with resolution (FWHM) $2.35\,\sigma_x$, then the expected number of events in a bin Δx around x_0 is

$$P = P_0 \int_{x_0 - \Delta x/2\sigma_x}^{x_0 + \Delta x/2\sigma_x} \frac{e^{-(x-x_0)^2/2\sigma_x^2}}{\sqrt{2\pi}\,\sigma_x} \, dx \qquad (7)$$

where P_0 is the total peak area. Thus, if A_α is the upper limit of the number of counts in the Δx region, the upper limit of the total peak area is

$$P_\alpha = \frac{A_\alpha}{P/P_0} . \qquad (8)$$

As can be seen from figure (1), A_α increases if both B and C increase. Since B and C increase with Δx then A_α also increases with Δx. On the other hand, the ratio P/P_0 increases asymptotically with Δx. Thus, $P\alpha$ has a minimum value as function of Δx. We will discuss below how to determine this optimum Δx bin in order to obtain the smallest P_α for a given C.L..

c) Uncertainty in the expected position of the searched peak

If x_0 is not exactly known and we know only an experimental result $\bar{x} \pm \sigma_{\bar{x}}$, this lack of information must be taken into account in the determination of the upper limit of the peak area. Adopting the Bayesion approach and supposing that the posterior

probability density function of x_0 is a Gaussian one with the parameters $\bar{x}$ and $\sigma_{\bar{x}}$ we can rewrite equation (7) as

$$P = P_0 \int_{-\infty}^{x} \left[\int_{\bar{x}-\Delta x/2\sigma_x}^{\bar{x}+\Delta x/2\sigma_x} \frac{e^{-(x-x_0)^2/2\sigma_x^2}}{\sqrt{2\pi}\,\sigma_x} \, dx \right] \frac{e^{-(x-\bar{x}_0)^2/2\sigma_{\bar{x}}^2}}{\sqrt{2\pi}\,\sigma_{\bar{x}}} \, dx_0 \quad . \tag{9}$$

In this equation the inner integral takes into account the fraction of the peak area in the Δx bin around $\bar{x}$ if the true position of the peak is x_0; the outer integral takes into account the probability that x_0 is the true position of the peak.

Integration in x_0 is readily done and equation (9) can be rewriten as

$$P = P_0 \int_{-\Delta x/2\sigma}^{\Delta x/2\sigma} \frac{e^{-u^2/2}}{\sqrt{2\pi}} \, du \tag{10}$$

where σ is given by

$$\sigma = \sqrt{\sigma_x^2 + \sigma_{\bar{x}}^2} \quad . \tag{11}$$

This last equation shows that both the peak standard deviation. σ_x, and the uncertainty in the peak position. $\sigma_{\bar{x}}$. have the same rule in the statistical problem.

III. OPTIMUM ENERGY BIN Δx

As stated above there is an optimum Δx bin to be analysed in order to obtain the smallest upper limit P_o. This optimum bin can be determined experimentally for every

case. However, as we are dealing with an hidden peak, we are able to generally determine Δx supposing that

$$B = C = k \sigma b \tag{12}$$

where b is the background rate (in counts per channel),

$$\Delta x = k \cdot \sigma \tag{13}$$

is the analysed bin, and

$$\sigma_B = k \cdot \sigma \cdot \sigma_b \tag{14}$$

is the standard deviation of B. σ_b is the standard deviation of b .

Using B and C from equation (12) and σ_B from equation (14) to obtain A_α from figure (1) and Δx from equation (13) to obtain P/P_0, we are able to obtain P_α from equation (8) as a function of k . Figure 2 shows how P_α depends on k in an example where b = 5 counts per channel, $\sigma_b/\sqrt{b} = 0.5$, $\sigma = 2$ channels and B = C. As can be seen from this example there is an optimum k value (and thus an optimum Δx bin) to be used.

Figure (3) shows $\bar{k}$, the optimum k value, as function of σ, b and $\sigma_b/\sqrt{b}$ calculated for $a = 0.05$ and supposing B = C . Some care must be taken in the use of figure (3) since when dealing with multichannel spectra we are limited to use integer Δx bins. So, $\bar{k}$ values from figure 3 are approximated values that must be rounded—off in order to obtain integer $k \cdot \alpha$ values. The assymptotic value of $\bar{k}$, when $\sigma_b = 0$ and $\sigma = \infty$, is 2.8 whatever b is. The dependence of $\bar{k}$ on a is very weak (only few percents if α goes from 0.7 to 0.01); thus, $\bar{k}$ from figure (3) can be used for a values that are not very far from 0.05.

IV. GAUSSIAN APPROXIMATION

In the following section we will show how to generally use the present results in the analysis of a hidden peak in a multichannel spectrum. Before this, we will show the Gaussian approximations valid in the case of not small number of events.

If $b \gg 1$ the posterior probability density function of a from equation (5) can be written as

$$g(a) = I' \cdot e^{-(a-\hat{a})^2/2\sigma_a^2}$$ (15)

where $I' '$ is a normalization constant

$$\hat{a} = C - B$$ (16)

and

$$\sigma_a^2 = C + (\hat{k} \sigma \sigma_b)^2 \quad .$$ (17)

In the case of a hidden peak, $B \cong C \cong \hat{k} b\sigma$ and $\hat{a} \cong 0$. A_α is given by

$$A_\alpha = \lambda_{\alpha/2} \sqrt{\hat{k} b \sigma + (\hat{k} \sigma_b \sigma)^2}$$ (18)

where

$$\frac{\alpha}{2} = \int_{\lambda_{\alpha/2}}^{\infty} \frac{e^{-\nu^2/2}}{\sqrt{2\pi}} d\nu \quad .$$ (19)

The first term in the square root of equation (18) corresponds to the statistical fluctuation of the number of events in the analysed Δx bin and the second term corresponds to the standard deviation of the mean value of the background. It is interesting to observe that both terms have the same rule in the determination of A_α.

When b is well known, $\sigma_b^2 \gg b$, $\xi = 2.5$. In this case P/P_0 can be readilly determined from equation (10) and thus, from equation (8), we have

$$P_a = 2.0\, \lambda_{a/2}\, \sqrt{b(\sigma_x^2 + \sigma_y^2)^{1/2}} \ . \tag{20}$$

where the dependence of σ on σ_x and σ_y from equation (11) was explicited.

In the appendix we compare the present result with that from ref.[4].

V. HOW TO PROCEED

Suppose we are searching for a Gaussian peak in a multichannel spectrum at $\bar{x} \pm \sigma_x$. Suppose that the system resolution (FWHM) is 2.35 σ_x. The following steps must be performed in order to determine the upper limit of the peak area:

a) Determine the background b , in counts per channel, and its standard deviation σ_b ;

b) Determine σ from equation (11) and use figure 3 to determine the best $\Delta x \simeq \xi\sigma$ region to be analysed. Δx must be rounded-off to an integer;

c) Determine $B = \Delta x \cdot b$ and $\sigma_B = \Delta x \cdot \sigma_b$. Determine also C . the number of counts in the bin Δx around $\bar{x}$;

d) Read from figure 1 the upper limit A_a ;

e) Determine P/P_0 from equation (10) using error function tables and use equation (8) to determine the upper limit of the peak area P_a .

As a pratical example, we can analyse the data from the $0^+ - 2^+$ ^{76}Ge neutrinoless double beta decay study of the Milano group [5]. This experiment consists in

the observation of a very low background Ge detector spectrum accumulated during 1.76y of running. The signature of the ^{76}Ge $0^+ \to 2^+$ transition is a peak at $Q_{\beta\beta} - E_\gamma$, where $Q_{\beta\beta}$ is the mass difference between ^{76}Ge and ^{76}Se (2041.4 ± 0.5 keV [6]) and E_γ is the energy of the first 2^+ level of ^{76}Se (559.11 ± 0.05 keV [7]), with $\sigma_z \simeq 2.3$ channels. The fiducial volume of the detector is 133 cm^3 [5]. From figure 14 of ref. [5] it was possible to evaluate the background from a 20 keV bin as 11.24 ± 0.46 counts per channel.

There is an important point to be discussed about σ_z: there are some different nuclear reactions linking ^{76}Ge to ^{76}Se allowing different experimental $Q_{\beta\beta}$ values for the $0^+ \to 2^+$ transition [6]. However, those ones with smaller standard deviations give imcompatible values, suggesting that probably systematic errors are present in the experiments. A constrained last square fit of the ^{76}Ge $- ^{76}$Se mass difference quoted by [6] gives a reduced chi-square value of about 12, suggesting that 0.5 keV is probably an understimation of σ_z. Thus, we will adopt here $\sigma_z = 1.6$ keV .

Using the above values, performing steps a) to e) and supposing as 50% the probability that deexcitation gamma-ray escapes from the detector, we obtain the following half-life limits to the $0^+ \to 2^+$ ^{76}Ge decay:

$$T_{1/2}(0^+ \to 2^+) > 0.91 \ 10^{22}y \quad (\text{at } 90\% \text{ C.L.})$$
$$> 0.76 \ 10^{22}y \quad (\text{at } 95\% \text{ C.L.}) \ . \tag{21}$$

In the case of 90% C.L. this result if about 1.7 times lower than the limit quoted by [5].

VI. DISCUSSION

The reason to adopt the Bayesian statistical approach, in this work, is not doctrinaire but practical. The Bayesian approach allows direct and easy answers to the problem of the difference between two Poissonian numbers without the necessity of approximations, as happens to the Classical approach [8]. The inclusion of non vanishing uncertainties both in the background and in the expected position of the searched peak is also direct. We also observe that in the case of experiments with small number of events the Bayesian approach seems to be closer to the expectation in experimental physics than the classical approach [9].

Considering qualitative aspects of the present procedure we can observe that:

a) The standard deviations of the background has the same rule of its statistical flutuation in the peak region (as can be clearly seen from equation (18)) in the determination of the upper limit of the number of counts in the analysed bin;

b) The standard deviation of the expected position of the searched peak has the same rule of the detector resolution;

c) The present approach can be used also if a peak – or a bump – is observed in the analysed region. In this case the upper limit of the peak area is compatible to the estimated peak area;

d) In any case the upper limit is negative or zero, even if there is a (statistical) depression in the analysed region.

e) The results can be applied in the case of small number of events.

APPENDIX

In ref. [4] the maximum likelihood technique was used in order to determine the upper limit of a Gaussian peak in a multichannel spectrum. That result can be compared with the result shown in this paper after the conection of a mistake in ref. [4] and some approximations.

Equation (6) of ref. [4] is better written as

$$\frac{1}{f_\lambda} \frac{df_\lambda}{dy_\lambda} \frac{dy_\lambda}{d\lambda} = \frac{1}{\sigma_y^2 \sqrt{2\pi} \sigma_x} \left[z_i - \frac{\lambda}{\sigma_x \sqrt{2\pi}} e^{-(x_i-x_0)^2/2\sigma_x^2} \right] e^{-(x_i-x_0)^2/2\sigma_x^2} , \qquad (A1)$$

where σ_x at the denominator of the first term in the right hand side of this equation was omitted in ref.[4]. Equation (9) and (10) of this reference must be rewritten respectivelly as

$$a' = \frac{1}{\sigma_y^2 \sqrt{2\pi} \sigma_x} \sum_{i=1}^{n} e^{-(x_i-x_0)^2/2\sigma_x^2} \qquad (A3)$$

and

$$b' = \frac{1}{2\pi \sigma_y^2 \sigma_x^2} \sum e^{-(x_i-x_0)^2/2\sigma_x^2} . \qquad (A4)$$

The meaning of the symbols of equations (A1) − (A3) are explained in [4].

As the peak is not seen, $a \gtrsim 0$. If σ_x is sufficiently larger than 1 and if the number of channels used in the analysis, n, includes all the peak region then

$$\sum_{i=1}^{n} \frac{e^{-(x_i-x_0)^2/2\sigma_x^2}}{\sqrt{\pi} \sigma_x} \simeq 1 . \qquad (A5)$$

Using this approximation the upper limit P_α from the approach of ref. [4] can be written as

$$P_\alpha = \lambda_{\alpha/2} \sqrt{2\sqrt{\pi}\ b\cdot\sigma_x} \qquad\qquad (A6)$$

where the meaning of b, σ_z and $\lambda_{\alpha/2}$ are explained in the text.
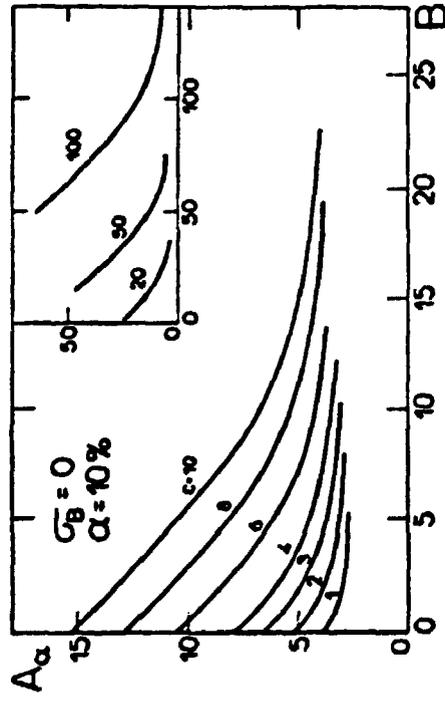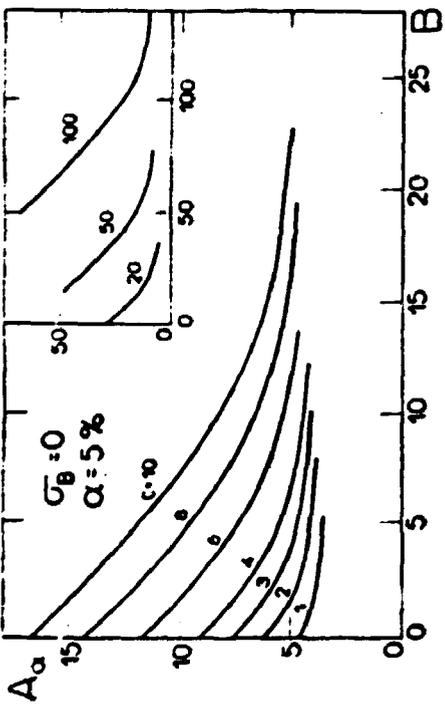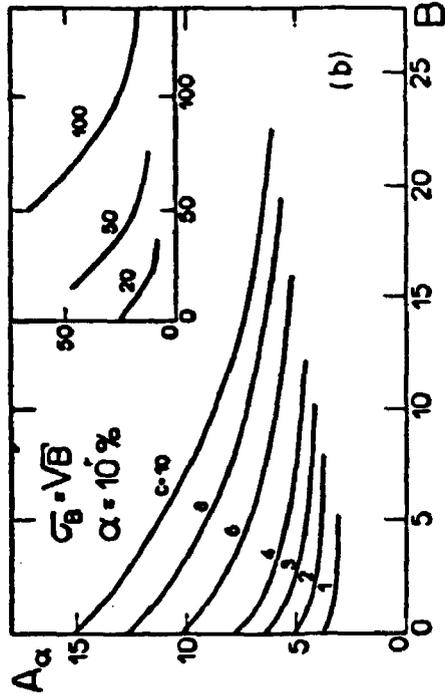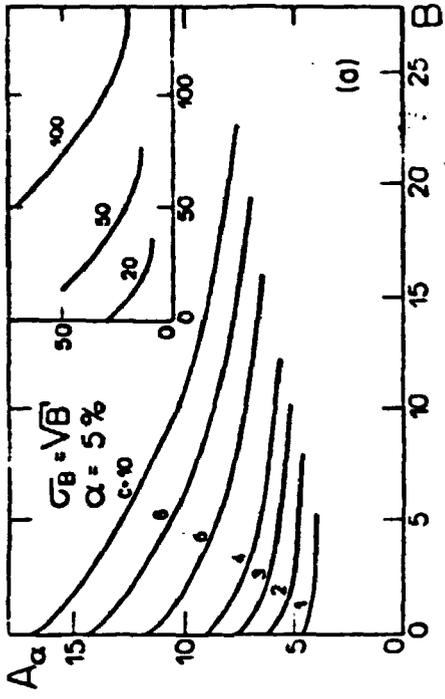
The result from ref. [4] can be compared with the result from this paper using the above equation (A5) and equation (20) with $\sigma_z = 0$. The numerical difference between both equations are about 6%.

REFERENCES

1) O. Helene, *Nucl. Instr. and Meth.* 212 (1983) 319.

2) M. Aguilar–Benitez et al., *Phys. Lett.* B204 (1988).

3) G. Zech, *Nucl. Instr. and Meth.* A277 (1989) 608.

4) F.T. Avignone III, *Nucl. Instr. and Meth.* A245 (1986) 525.

5) E. Bellotti et al., *Nuovo Cimento* 95A (1986) 1.

6) R.J. Ellis et al., *Nucl. Phys.* A435 (1985) 14.

7) C.M. Lederer and V.S. Shirley, *Table of Isotopes*, 7th edition (Willey, New York, N.Y., 1978).

8) H.B. Prosper, *Nucl. Instr. and Meth.* A238 (1985) 500.

9) B. Escoubes et al., *Nucl. Instr. and Meth.* A257 (1987) 346.

16

FIGURE CAPTIONS

Figure 1.   Upper limit A_α as a function of B for different C values for (a) 95% C.L. and (b) 90% C.L.. In both cases A_α was determined for $\sigma_B = 0$ and $\sigma_B = \sqrt{B}$.

Figure 2. P_α, $\alpha = 0.05$, as a function of k for a hidden peak with $\sigma = 2$ channels. The optimum k value is about 2.2 which corresponds to $\Delta x = 4.4$ channels. Approximating this values to $\Delta x = 5$ or $\Delta x = 4$ has not a very important consequence. However, if a bad Δx region is choosen then the upper limit of the total peak area will be significantly greater than the better one.

Figure 3. $\bar{k}$ values are shown for different values of b, σ and $\sigma_b/\sqrt{b}$. The asymptotic limits of $\bar{k}$ for $b \to \infty$ are shown in small symbols at the right side of the values for b = 100.
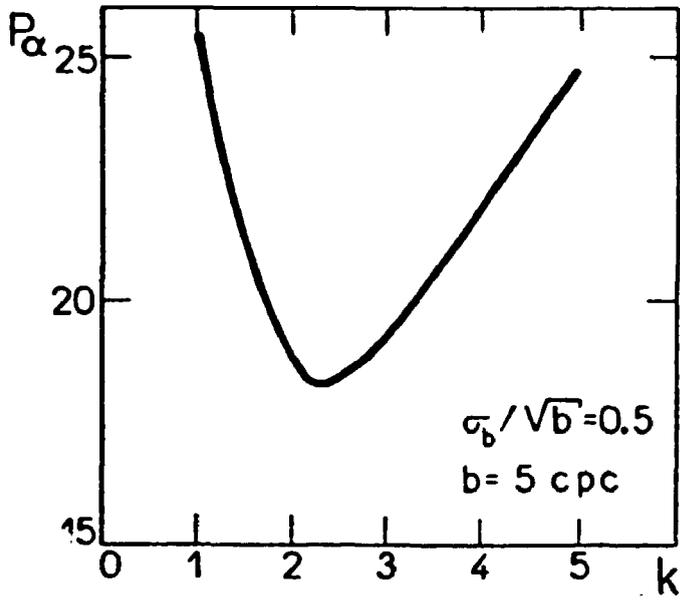
$\sigma_B = \sqrt{B}$
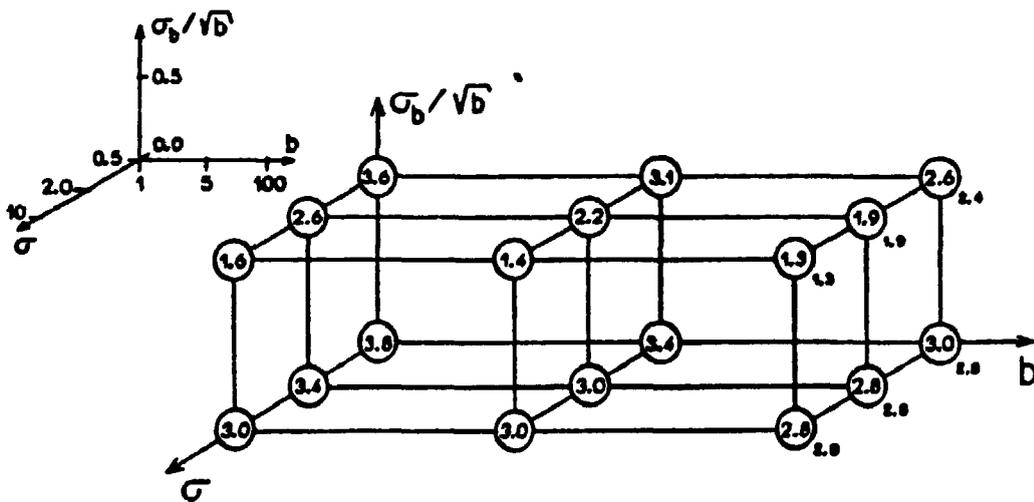$\alpha = 5\%$
A_α
B
(a)
$c = 10$
8
6
4
3
2
1
100
50
20

$\sigma_B = \sqrt{B}$
$\alpha = 10\%$
A_α
B
(b)
$c = 10$
8
6
4
3
2
1
100
50
20

$\sigma_B = 0$
$\alpha = 5\%$
A_α
B
$c = 10$
8
6
4
3
2
1
100
50
20

$\sigma_B = 0$
$\alpha = 10\%$
A_α
B
$c = 10$
8
6
4
3
2
1
100
50
20

FIGURE 2

FIGURE 3