UNIVERSIDADE DE SÃO PAULO

INSTITUTO DE FÍSICA
CAIXA POSTAL 20516
01000 - SÃO PAULO - SP
BRASIL

publicações

IFUSP/P-354

UPPER LIMIT OF PEAK AREA

by

O. Helene

Instituto de Física, Universidade de São Paulo,
São Paulo, Brasil

Agosto/1982

UPPER LIMIT OF PEAK AREA*

O. Helene

Instituto de Física, Universidade de São Paulo,
São Paulo, Brasil

RESUMO

Discutimos aqui a determinação do limite superior para a área de um pico em multicanal, com nível de significância conhecido. Esta questão é particularmente relevante nos casos em que a presença do pico é camuflada pela flutuação estatística do fundo. Os cálculos são feitos exatamente, permitindo que os resultados sejam aplicados nos casos de baixa estatística, quando não são válidas as aproximações gaussianas. Os resultados são comparados com simulações pelo método de Monte Carlo e aplicados no caso de decaimento beta do ^{92}Nb .

ABSTRACT

We discuss here the determination of the upper limit of peak area in a multi-channel spectra, with a known significance level. This problem is specially important when the peak area is masked by the background statistical fluctuations. The problem is exactly solved and, thus, the results are valid in experiments with small number of events. The results are submitted to a Monte Carlo test and applied to the ^{92}Nb beta decay.

* Submitted to Nuclear Instruments and Methods.

1. INTRODUCTION

A very frequent problem in spectroscopy is the non identification of a sought after peak in a multi-channel spectra. In this situation we cannot decide for a non-existing peak. We can only say that the peak area is less than or comparable to the background statistical fluctuation. However, what is the correct value to the upper limit of the area and the agreeable significance level? We discuss here this question, starting from usual hypothesis in statistical methods in experimental physics.

The upper limit of the peak area, A , must depend on the experimental results: the background and total counts in the peak region. The total counts in the peak region will be denoted by C . The background will be denoted by B , if its true value is known (or its standard deviation is negligible). Firstly, we will suppose that this hypothesis is true and subsequently we will extend the calculations to situations where the background is not exactly known. A must depends also on the desired significance level, identified by $(1-\alpha).100\%$. What we hope is to say that the peak area is less than A , being α the error probability of the affirmation.

The statistical treatment presented is not unique for peak area in multi-channel spectra. It is applicable to any equivalent experiment involving a statistical background and a contribution due to a source.

In our approach the first and fundamental step is the determination of the probability density function (p.d.f.) of the peak area after the experiment. This p.d.f. will be identified by $g(a)$. The variable a is the peak area, if we consider it as a statistical variable[1,2]. In a more rigorous formalism, $g(a)$ is the degree of belief in different possible values of a[3]. The p.d.f. of the area after the experiment should depends on the results, B and C, and on the prior p.d.f. . The second step is the determination of A, which can be easily made if we know $g(a)$.

This same problem, the determination of upper limits for non observed peaks, has been discussed by some authors, with a different approach[4,5,6]. Those authors consider two hypothesis: the peak exist and the peak does not exist. If the observed count exceed a critical limit A_I, they recommend to decide for a non zero area; if this critical value is not exceeded, they recommend to decide for a zero area (the peak does not exist)[6]. Nevertheless, there are some problems in that approach. If the observed count exceed slightly the critical value A_I, we should decide for a finite (non zero) area; if the count is slightly below the critical value, we should decide for a zero area. Thus, there is a discontinuity when the count is equals to the critical value. This problem is due to the two initial hypotheses, the peak exists and the peak does not exist, which are discreets. Other problem is in the decision "the peak does not exist". This is always a impossible decision or, likewise, it is a decision with a zero significance level. We can never decide for an exact value of a measured quantity, whatever experiment we are dealing. This implies a zero standard deviation and, thus, is impossible.

Add to that, other problems related to that approach has been pointed[7].

2. PROBABILITY DENSITY FUNCTION OF PEAK AREA

If the mean contributions of the peak area and of the background in a multi-channel region are a and B respectively, the probability of one obtaining C counts is

$$P(C) = \frac{e^{-(a+B)}(a+B)^C}{C!} \qquad (1)$$

This is a Poisson distribution with mean $a+B$, and is correct if a and B are the true values. In eq. (1) a and B are parameters and C a discrete variable.

The inverse problem, the determination of the p.d.f. of a after the experiment, was stablished very early[1,2]: if the p.d.f. of a before the experiment is a uniform frequency function on the axis $(0,\infty)$, them the right side of eq. (1) is proportional to the p.d.f. of a after the measurement. This result can be explicitly shown using the Bayes theorem[3]. Thus,

$$g(a) = N_1 \frac{e^{-(a+B)}(a+B)^C}{C!} , \qquad (2)$$

where N_1 is a normalization constant such that

$$\int_0^\infty g(a)\, da = 1 . \qquad (3)$$

In eq. (2) B and C are parameter and a a continuous variable.

There is a usual gaussian approximation of eq. (2),

$$g(a) \cong N_2 \frac{e^{-(a-\bar{a})^2/2C}}{\sqrt{2\pi C}} , \qquad (4)$$

where $\bar{a} = C-B$ and N_2 is a normalization constant. This approximation is valid if $C >> 1$, and is familiar for spectroscopists: the estimated value for the peak area is $\bar{a}$ and $\sqrt{C}$ its standard deviation.

Equations (2) and (4) are valid when the background is exactly known or has a negligible standard deviation. If this does not occur, the expression for $g(a)$ is

$$g(a) = N_3 \int \frac{e^{-(a+B)}(a+B)^C}{C!} f(B)\, dB \quad . \tag{5}$$

In this equation N_3 is a normalization constant and $f(B)$ the background p.d.f. Frequently $f(B)$ has a gaussian shape with mean B_0 and standard deviation σ_B. If so, and if $C >> 1$, we can expand the integrand to obtain an approximate expression for $g(a)$,

$$g(a) \cong N_3 \frac{e^{-(a-\bar{a})^2/2\sigma^2}}{\sqrt{2\pi}\,\sigma} \quad . \tag{6}$$

In this equation $\bar{a} = C-B$ and $\sigma^2 = \sigma_B^2 + C$. Equation (6), as equation (4), is usually adopted, being $\bar{a}$ the estimate of the peak area and σ its standard deviation.

3. UPPER LIMIT OF PEAK AREA

If we know $g(a)$ we can determine the upper limit to the peak area, with a desired significance level. The probability of an a value greather than A is

$$\alpha = \int_A^\infty g(a)\, da \quad . \tag{7}$$

The probability of an a value less than A is $1-\alpha$. Thus, we can say that the peak area is less than A, with a significance level $(1-\alpha).100\%$. The A value depends on the experimental results, B and C if the error in the background is negligible, or B_0, σ_B and C, if the background standard deviation is σ_B, and on α.

Figures 1 and 2 show plots os A versus B for some C values. These figures were obtained using equation (2) in equation (7) and are valid if the background standard deviation is negligible. For example, if an experiment $B = 5.3$ and $C = 4$, we can say that the peak area is less than $A = 5.3$ with 95% significance level ($\alpha = 0.05$). If, at the same background condition, $B = 5.3$, $C = 6$, the A value is 7.1, with the same (95%) significance level.

An important property of the upper limit is the addition property. If both examples cited above, $B = 5.3$ and $C = 4$ and $B = 5.3$ and $C = 6$, correspond to the same peak in two experiments, we must add both to obtain the over-all upper limit. Thus, we analyse the result $B = 10.6$ and $C = 10$. This result corresponds to $A = 7.8$ with 95% confidence level. This addition property can be proved using, in the Bayes theorem[3], the p.d.f. of a after the first experiment as the prior p.d.f. of a in the second experiment. We can demonstrate it also when the background is not the same and for more than two experiments.

It is interesting to note that when $C = 0$, the p.d.f. of a after the experiment is

$$g(a) = e^{-a} \quad , \tag{8}$$

whichever is the B value. The upper limits to the peak area in this case are 3,00 and 2,30 , respectively for $\alpha = 0.05$ and $\alpha = 0.10$.

If we use the gaussian approximation equations (4) or (6), in expression (7), we obtain a comfortable expression for α. Using the error function

$$I(z) = \frac{1}{\sqrt{2\pi}} \int_z^\infty e^{-x^2/2}\, dx \quad , \tag{9}$$

we obtain

$$\alpha = \frac{I\left(\frac{A-\bar{a}}{\sigma}\right)}{I\left(\frac{-\bar{a}}{\sigma}\right)} \quad . \tag{10}$$

In this equation $\bar{a} = C-B$ and $\sigma = \sqrt{C}$, if the background standard deviation is negligible, or $\bar{a} = C-B_0$ and $\sigma^2 = C+\sigma_B^2$, if the standard deviation of the background is σ_B . The addition property discussed above is also valid in this case of gaussian approximation.

We must observe that, no matter whichever are the experimental results, we never have $A = 0$. The value $A = 0$ corresponds to $\alpha = 1$ in equation (7) and, thus, to a zero significance level. Consequently, we never decide for a non existing peak.

4. A MONTE CARLO TEST

The basic hypothesis in our approach is the prior p.d.f. of a . Before the experiment we suppose a total ignorance about the peak area and this is quantified adopting a uniform frequency function in the range $(0,\infty)$. Thus, in a Monte Carlo simulation, we need to generate a true "unknown" a value in this range. Nevertheless, this is impossible but, luckily, unnecessary. We can restringe the upper limit in a W value, if the probability that an a value greather than W generate C counts is negligible.

In the Monte Carlo test it was set the B and C values and generated the true "unknown" a value, uniformly distributed in the range $(0,W)$. After that it was generated a count value, obeying a Poisson distribution with mean value $a+B$. If this count value was not equals to C , the a value was discarded and a new one generated. If the count value was equal to C the a value was considered. Following, it was examined if the a value is or is not less than the upper limits. Table 1 shows the results of 1,000 a values considered. The last two columns show the total number of incorrect decisions about the true a value. The expected numbers in these two columns are 50 and 100 respectively. The differences between these expected values and the observed are compatible with the statistical fluctuations.

The results of Table 1 correspond to low C values and to a negligible standard deviation of the background.

Table 2 shows a Monte Carlo test for $C >> 1$ and a non-zero background standard deviation. In this simulation the background was normaly generated, with mean B_0 and standard deviation $\sigma_B = \sqrt{B_0}$. After that, it is generated a true "unknown" a value and then a counting value. If this counting was equal to a initially choosen C value, the experiment was considered;

if not, a new background and a new true "unknown" peak area was generated. The upper limits of Table 2 were calculated using equation (10). It was simulated 500 experiments.
The differences between expected and observed values to erroneous prevision of the peak area are not only because statistical fluctuations but also because the involved approximations.

5. THE ^{92m}Nb DECAY

We measured the residual gamma-ray activity that follows the ^{92}Nb beta decay. The purpose was determine the feeding of the 1383 keV, 1496 keV and 2067 keV levels of ^{92}Zr. The feeding of this last level was determined as $5.2 \cdot 10^{-3}\%$[8]. The feeding of the first two, characterized by 449 keV and 562 keV gamma-ray transitions[8,9], was not observed. The knowledge of the total counts and of the background in the peak regions can be used to determine upper limits to the peak areas. These, by their turn, allow the determination of upper limits to the beta transitions to 1383 keV and 1496 keV ^{92}Zr levels.

The ^{92}Nb source was produced in ^{93}Nb (γ,n) reaction. The source was a metallic niobium layer, 1.0 g/cm^2 thick. It was made three countings of the residual gamma-ray activity which were summed. It was used a 53 cm^3 Ge(Li) detector and an Ortec 572 amplifier with pile-up rejector. The system resolution was about 2.5 keV at 1173 keV. Ref. 8 describes some others experimental details.

Table 3 shows the obtained results. The C values correspond to the total counts in a 8 keV region centered on the peak. The B_0 values correspond to the total counts of two

4 keV regions at the peak region neighbourhood. We assume that these are the mean background value in the peak region, with standard deviation $\sqrt{B_0}$. Using equation (10) we determine upper limits to the 449 keV and 562 keV peak areas, for $\alpha = 0.05$ (95% confidence level). Knowing upper limits to peak areas, conversion electrons coefficients[9,10], absorption coefficients[11], and using the 934 keV gamma-ray transition as reference[9], we can determine upper limits to the beta transitions.

The upper limits of ^{92m}Nb decay to 1383 keV and 1496 keV ^{92}Zr levels are $3.3 \cdot 10^{-3}\%$ and $4.5 \cdot 10^{-3}\%$ respectively, with 95% significance levels.

The last column of Table 3 shows the lower limits of log ft values, determined using usual tables and nomograms[12]. These values accords to second forbidden transitions, as expected by spin and parity assignements[9].

REFERENCES

1) W.L. Nicholson, Nucleonics 24, Nº 8 (1966) 118
2) L.J. Rainwater and C.S. Wu, Nucleonics 1, Nº 2 (1947) 60
3) W.T. Eadie, D. Drijard, F.E. James, M. Ross and B. Sadoulet, "Statistical Methods in Experimental Physics", North-Holland Publ. Co., Amsterdam, 1971.
4) B.S. Pasternack and N.H. Harley, Nucl. Instr. and Meth. 91 (1971) 533
5) B. Altshuler and B. Pasternack, Health Phys. 9 (1963) 293
6) J.J. Donn and R.L. Wolke, Health Phys. 32 (1977) 1.
7) B.J. Barry, Health Phys. 36 (1979) 208
8) O. Helene, Thesis, Instituto de Física da Universidade de São Paulo, São Paulo, Brazil, 1982
9) P. Luksch, Nucl. Data Sheets 30 (1980) 573
10) F. Rosel, H.M. Fries, K. Alder and H.C. Pauli, Atom. and Nucl. Data Tables, 21-2 (1978) 92
11) C.M. Davisson, in "Alpha-, Beta-, and Gamma-Ray Spectroscopy", edited by K. Siegbahn, North-Holland Publ. Co., Amsterdam, 1965.
12) C.M. Lederer and V.S. Shirley, "Table of Isotopes", 7ª ed., John Wiley & Sons, Inc., 1978.

C	B	W	$A_{0.05}$	$A_{0.10}$	values greather than $A_{0.05}$	values greather than $A_{0.10}$
0	3	10	3.0	2.3	50	106
8	5	24	9.6	8.2	43	95
10	5	26	12.0	10.4	45	99
10	10	22	8.1	6.6	45	103

TABLE 1 - Results of the Monte Carlo test for an exact knowledge of B. The upper limits, $A_{0.05}$ and $A_{0.10}$, were determined using eq. (2) in eq. (7).

B_o	C	W	$A_{0.05}$	$A_{0.10}$	values greather than $A_{0.05}$	values greather than $A_{0.10}$
50	50	50	19.6	16.4	35	58

TABLE 2 - Results of the Monte Carlo test supposing a background with standard deviation $\sqrt{B_o}$. The upper limits $A_{0.05}$ and $A_{0.10}$ was determined using eq. (10) with α=0.05 and α=0.10, respectively.

E_γ(keV)	C (10^3)	B_o (10^3)	$A_{0.05}$ (10^3)	I_t $(10^{-3}\%)$	log ft
449	7467.4	7467.1	7.8	< 3.3	> 10.2
561	7909.6	7908.7	8.5	< 4.5	> 9.9

TABLE 3 - Upper limit to 449 keV and 561 keV gamma-ray peak areas ($A_{0.05}$) in ^{92m}Nb decay and upper limit to beta transitions (I_t). The results correpond to 95% confidence level.

FIGURE CAPTIONS

FIGURE 1 - A versus B for some C values to 95% ($\alpha = 0.05$) confidence level.

FIGURE 2 - The same of Figure 1, here to 90% ($\alpha = 0.10$) confidence level.

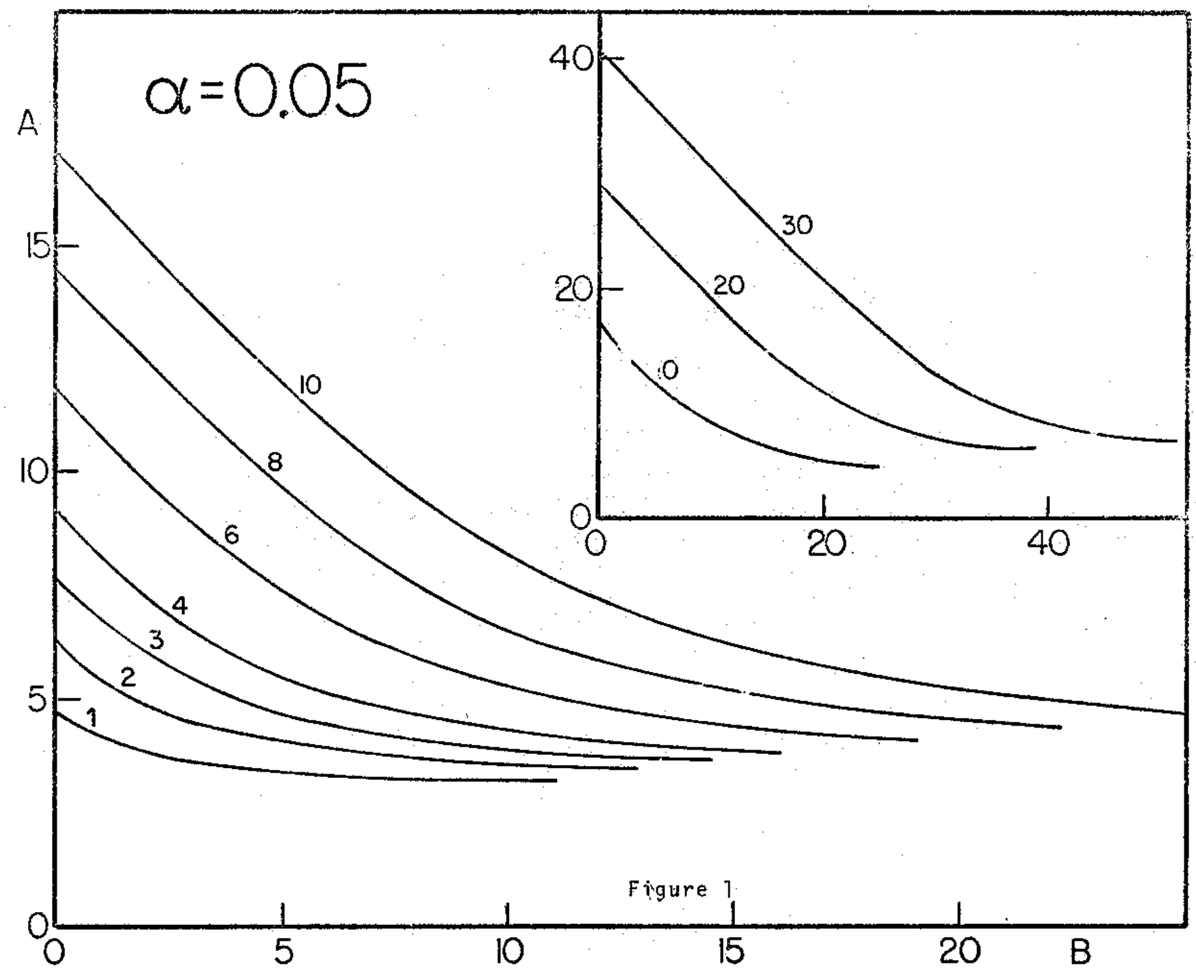

Figure 1

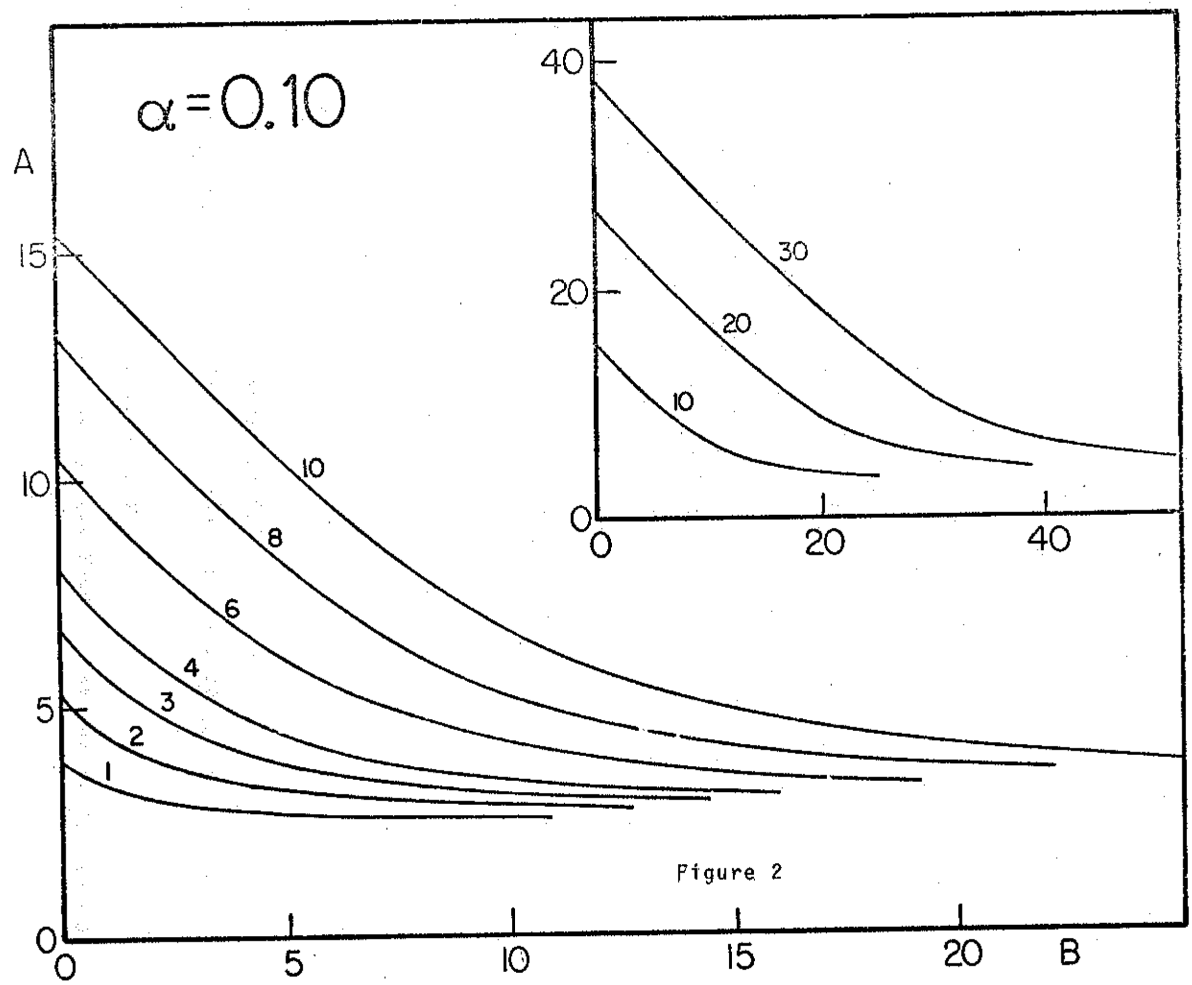

Figure 2