STATISTICAL TREATMENT OF NUCLEAR COUNTING RESULTS

by

Čemal B. Dolićanin 1*, Koviljka Dj. Stanković 2, Diana Č. Dolićanin 1, and Boris B. Lončar 3

1State University of Novi Pazar, Novi Pazar, Serbia
2Faculty of Electrical Engineering, University of Belgrade, Belgrade, Serbia
3Faculty of Technology and Metallurgy, University of Belgrade, Belgrade, Serbia

Since the exact time a specific nucleus undergoes radioactive decay cannot be specified, nor can showers caused by secondary cosmic rays be predicted, statistical laws play an important role in almost all cases of experimental nuclear physics. This paper describes the method for the statistical treatment of nuclear counting results obtained experimentally by taking into account random variables pertaining to both frequent and infrequent phenomena. When processing counting measurement data, it is recommended to first discard spurious random variables that spoil the statistics by using Chauvenet’s criterion, as well as to test if the results in the statistical sample follow a unique statistical distribution by using the Wilcoxon rank-sum test (U-test). The verification of the suggested statistical method was performed on counting statistics obtained both from the radioactive source Cs-137 and background radiation, expected to follow the normal distribution and the Poisson distribution, respectively. Results show that the application of the proposed statistical method excludes random fluctuations of the radioactive source or of the background radiation from the total statistical sample, as well as possible inadequacies in the experimental set-up and show an extremely effective agreement of the theoretical distribution of random variables with the corresponding experimentally obtained random variables.

Key words: counting statistics, Chauvenet’s criterion, U-test

INTRODUCTION

By observing natural radioactivity, it was concluded that it consists in a delayed spontaneous transformation of the nucleus accompanied by the emission of radiation (particles and/or photons). The exact time when a specific nucleus will transform, i.e., undergo radioactive decay, cannot be specified. It is, however, independent of the previous history of the nucleus, i.e., there is no memory effect. The half-life of a nucleus is, therefore, introduced as a characteristic of this phenomenon, defined as the mean decay time around which the decay times of a large number of the same type of nuclei fluctuate [1, 2]. Radioactive decay itself is a consequence of the fact that the initial nucleus is more energetically unstable than the resulting one.

For this reason, in almost all cases of experimental nuclear physics, statistical laws play an important role. This is most pronounced in the processing of measurement results obtained from counters, i.e., when the results of nuclear counting are treated. It is usually assumed that nuclear counting results follow either the normal distribution, in case the statistical sample consists of random variables pertaining to a phenomenon that occurs many times during a unit of time, or the Poisson distribution, in case the statistical sample contains random variables referring to an infrequent phenomenon.

However, this kind of statistical treatment is oversimplified, since it ignores fluctuations of the considered nuclear phenomena. These fluctuations are present even in the case when the statistical sample consists of random variables describing a frequently occurring phenomenon, e.g., during spectroscopic measurements with radioactive sources, but are of greater importance when the statistical sample contains random variables referring to an infrequent phenomenon, e.g., in the case of background radiation measurements, when sudden showers of counter impulses caused by secondary cosmic rays may appear. Furthermore, the described oversimplified treatment is inadequate when there is a possibility that the statistical distribution of nuclear counting results is of a complex type, either additive or multiplicative.

* Corresponding author; e-mail: rektorat@np.ac.rs
It is therefore advisable, when counting results are processed, to first discard spurious random variables that spoil the statistics, as well as to test if the results in the statistical sample follow a unique statistical distribution. The aim of this paper is to propose such additional methods for treating nuclear counting results.

**CHAUVENET’S CRITERION AND WILCOXON RANK-SUM TEST (U-TEST)**

The Chauvenet’s criterion, formulated a long time ago as a means for rejecting erroneous readings of star locations with a sextant, can be used to discard the majority of spurious random variables arising in an experimental procedure. The Wilcoxon rank-sum test (i.e., the U-test) is an algorithm for establishing whether all statistical samples, obtained by dividing a larger sample, follow a unique statistical distribution. By applying these two methods, it is possible to reject either specific random variables, or subgroups of random variables that spoil the quality of the statistical treatment and the reliability of the obtained results [3, 4].

Chauvenet’s criterion

The problem of determining the value of a physical quantity from the repeated measurements of parameters performed as a part of an experiment is linked to the general analysis of the realization of experiments or measurements, i.e., to the statistical analysis of the obtained results. The set of all results obtained during one or several measurement series is taken for a population, while certain subgroups of results from one or several measurement series form the samples

\[
\begin{align*}
    x_1, x_2, \ldots, x_j, \ldots, x_k, \ldots, x_N & \quad \text{– population} \\
    x_1, x_2, \ldots, x_j, \ldots, x_n & \quad \text{– a sample (} n < N \text{)} \\
    x_1, x_{i+1}, \ldots, x_j, \ldots, x_k & \quad \text{– a sample (} k + 1 < N \text{)}
\end{align*}
\]

After the general analyses of the experiment itself, the first task is to calculate the mean value of the measured quantities over the whole population and over the samples \((x_p, x_s)\), as well as the mean-square deviation over these same groups of results \((\sigma_p, \sigma_s)\). It is also useful to calculate the mean-square deviation of the mean values for all investigated quantities \((\sigma_{xp}, \sigma_{xs})\)

\[
\begin{align*}
    x_p &= \frac{\sum_{i=1}^{n} x_i}{N} \quad ; \quad \sigma_p = \sqrt{\frac{\sum_{i=1}^{n} (x_i - x_p)^2}{N}} \\
    x_s &= \frac{\sum_{i=1}^{n} x_i}{n} \quad ; \quad \sigma_s = \sqrt{\frac{\sum_{i=1}^{n} (x_i - x_s)^2}{n-1}}
\end{align*}
\]

\[\sigma_{xp} = \frac{\sigma_p}{\sqrt{N}} \quad ; \quad \sigma_{xs} = \frac{\sigma_s}{\sqrt{n}} \] (4)

In recent literature, the mean-square deviation of the measured quantities \((\sigma_p, \sigma_s)\) is termed standard uncertainty of individual measurements, while the mean square deviation of means \((\sigma_{xp}, \sigma_{xs})\) is termed type A standard uncertainty [4].

The spectrum of measurement results

Analysis of the results obtained by performing measurements on the population and on different samples from that population, after mean values and mean square deviations have been calculated, consisted in sorting the sets of measurement results to which these calculations pertain. All the results constituting the population and the samples are sorted in an ascending and descending order. The sorted sets of values then clearly show the positions of the calculated mean values, as well as the coverage of measurement results by the respective mean square deviations

\[
\begin{align*}
    x_{\min} & \leq x_1 \leq x_p - \sigma_p \leq \ldots \leq x_j \leq x_p \leq \ldots \\
    \ldots & \leq x_N \leq x_p + \sigma_p \leq \ldots \leq x_{\max} \\
    x_{\min} & \leq x_1 - \sigma_s \leq \ldots \leq x_k \leq x_1 \leq \ldots \\
    \ldots & \leq x_M \leq x_s + \sigma_s \leq \ldots \leq x_{\max}
\end{align*}
\] (5)

It can easily be detected how large the deviations of the limiting values of a set are relative to the mean value of the specific sorted set of results. Preliminary assessments on whether the limiting values should be treated separately can then be made. In most analyses, special attention is paid to treating the minimum and the maximum measured value, primarily regarding probabilities for their appearance [5-7].

Defining the limiting probabilities

The four basic calculated quantities (mean value, mean square deviation, maximum value, and minimum value) enable the determination of an interval that encompasses all measurement results, or just one part of the results relative to the population or the sample. An interval containing the results is expressed as a multiple of the mean square deviation \((\pm k \sigma, +k \sigma)\), or indirectly, by using parameters of a Gaussian distribution.

In general, all results are possible, meaning that the interval \((\pm k \sigma, +k \sigma)\) is infinite. However, during result analysis, a limiting value that is deemed to encompass values arising from fluctuations inherent in the measurement process is often set, while results outside that interval demand special consideration. The determination of this limiting value, or of the limiting interval width, is accomplished on the basis of an assumed or derived probability for a specific result to
appear as expected, moderately expected, or unexpected.

The said conclusion, that the probability of a result being either expected or unexpected gives rise to a numerical calculation of interval limits, makes the basis for our starting assumption.

The starting assumption is that event $E$, which is considered rare, to occur within $n$ executions of the experiment, i.e., measurements. The probability for event $E$ to occur in $n$ trials, with $n$ being a large number, is $1/n$, while the probability for event $E$ not to occur is $1 - 1/n$. The probability for event $E$ not to occur in $n$ trials is $(1 - 1/n)^n$. The probability for event $E$ to occur at least once in $n$ trials is

$$P(E \text{ occurs once or several times}) = 1 - \left(1 - \frac{1}{n}\right)^n$$

Since, $\lim(1-1/n)^n = \lim[(1-1/n)^n]^{-1} = e^{-1}$, it follows that $P(E) = 1 - \left(1 - \frac{1}{n}\right)^n = 1 - 1/e = 0.632$.

It is evident that for $n$ executions of an experiment, with $n$ being a large number, an unexpected event has a pretty high probability of occurring [8, 9].

The criterion for determining limiting probabilities (Chauvenet’s criterion)

We will suppose that $n$ measurements have been performed, with $n$ being a large number, and that a low-probability result has appeared. As a result of random fluctuations during measurements, it can be expected that the probability for any of the $n$ different results to appear is not much lower than $1/n$. If $n$ is a large number, $1/n$ is also the small probability for an unexpected or rare event $E$ to occur. Let the minimum probability for the unexpected event $E$ (i.e., the unexpected measurement result) to occur be defined as one half of the probability for one of $n$ different results to appear, with $n$ being large ($1/2n$). From this assumption, it follows that the probability for the unexpected event not to occur is

$$P = 1 - \frac{1}{2n}$$

and if, additionally, the distribution that describes the measurement process is Gaussian, the following relation ensues

$$P(t) = \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt = 1 - \frac{1}{2n}$$

From relation (8), it can thus be concluded that the value of parameter $t$ is

$$t = F^{-1}\left(\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-t^2/2} dt = 1 - \frac{1}{2n}\right) = F^{-1}(n) \tag{9}$$

Since parameter $t$ is defined as $t = (x_s - x)/\sigma_s$, it follows that

$$t = F^{-1}(n) = t(n) q_s = \frac{|x_s - x|_{\text{max}}}{\sigma_s} = e_{\text{max}}$$

which is the analytical form of the Chauvenet’s criterion. It is evident that $e_{\text{max}} = q_s t$, i.e., $e_{\text{max}} = t \sigma_s$. Let it be noticed that if the probability for a rare event is assumed to be $1/2n$, where $n$ is a large number, then the probability for event $E$ to occur at least once in $n$ trials is

$$P(E) = 1 - \left(1 - \frac{1}{2n}\right)^n$$

$$= 1 - \left[1 - \frac{1}{2n}\right]^{2n} = 1 - e^{-t/2} = 0.393 \tag{11}$$

meaning that an “unexpected event” is not too improbable, i.e., that measurement results lying outside the interval $q_s = t = F^{-1}(n)$ are considered “rare” and submitted to special analysis within their respective population or sample.

The operating procedure for implementing Chauvenet’s criterion

The operating procedure for implementing the Chauvenet’s criterion consists of several steps. It is first checked if the population distribution is Gaussian or is assumed to be Gaussian, if this is obvious. The group parameter $q_n$ (Chauvenet’s parameter) is then determined as

$$q_n = F^{-1}(n) = F^{-1}\left(\frac{2}{\sqrt{2\pi}} \int_{-\infty}^{t/2} e^{-x^2/2} dx = 1 - \frac{1}{2n}\right) = t \tag{12}$$

The table containing Chauvenet’s parameters for a number of different measurements $n$ is provided in ref. [10]. In the next step, for the set of results $x_1, x_2, ..., x_n$, the mean and the standard deviation are determined as

$$\bar{x} = \frac{\sum x_i}{n}; \quad \sigma = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n - 1}} \tag{13}$$

Next, parameters $q_i$ are determined as

$$q_1 = \frac{|x_1 - x|}{\sigma_1}; \quad ..., q_i = \frac{|x_i - x|}{\sigma_i}; \quad ..., q_n = \frac{|x_n - x|}{\sigma_n} \tag{14}$$

In the end, the set of Chauvenet’s parameters is sorted and compared to the group parameter $q_n$. If any of the parameters $q_i$ is larger than $q_n$, then parameter, or the result corresponding to it, is separated from the population or sample (i.e., eliminated or treated separately), and the procedure repeated.

Wilcoxon rank-sum test

One of the most important problems of mathematical statistics is the testing of the hypothesis about the equality of two distributions based on two samples taken from them, or the problem of testing the hypothe-
The rank-sum test. The equal number of elements, any of the alternative hypotheses to respective samples. The elements of the null hypothesis, $H_0$, can be formulated as “no difference between these two sets”, meaning that the two independent random samples belong to identical basic sets. It can be seen from fig. 1 that, if the null hypothesis is true, the two distributions coincide completely. If, on the other hand, the alternative hypothesis $H_1$ is true ($M_1 < M_2$), set B has a greater median and is shifted to the right of set A by $\Delta$.

The testing procedure is rather simple, founded on the ranking of joint observations from both samples, whereby it is important to keep track of the affiliation of values to respective samples. The elements of both sets are sorted according to size, so that rank 1 is assigned to the smallest of $n - n_1 + n_2$ combined observations, rank 2 to the next larger observation, etc., while rank $n$ is assigned to the largest datum. If there are data with equal values, their mean rank is found (e. g. if the sixth and the seventh element have equal values, both are assigned rank $(6 + 7)/2 = 6.5$). Test statistics is obtained as the sum of ranks in a sample, in the manner described next. Let the arrays of data in the samples, sorted according to size, be presented in the same graph by different symbols (e. g. a dot and a square), denoting the affiliation to respective samples (fig. 2).

If the null hypothesis that both samples originate from the same basic set is true, then the joint observations from the samples (as well as their ranks) will be completely mixed, as in fig. 2(a). If, however, observations with higher values (and higher ranks) appear much more often in the B set sample, as in fig. 2(b), it can be concluded that population B is shifted to the right with respect to population A, i. e. that it has a larger median. Since higher ranks appear mostly in B set, their sum can be taken for the test statistics because it, too, will be much higher than the rank sum of the population of the A sample. This would, therefore, indicate that the null hypothesis about the equality of the two samples’ medians should be rejected, i. e. these two samples could not have belonged to the same basic set [10].

The testing procedure

Let us consider again sets A and B with $n_1$ and $n_2$ elements, respectively. In order to make the use of critical value tables more convenient, we will introduce a convention that if the two samples differ in size, $n_1$ represents the number of elements in the smaller sample (set A), while $n_2$ is the number of elements in the larger sample (set B). The statistics of the rank-sum test $W_A$ is equal to the sum of ranks in the smaller sample. If the samples have an equal number of elements, any of them can be chosen for the calculation of the test statistics. Similarly, $W_B$ is obtained as the sum of ranks of the elements that belong to the larger sample (set B). Seeing that the sum of $n$ successive natural numbers is $n(n + 1)/2$, the sum of the test statistics $W_A$ with the rank sum $W_B$ from the sample with $n_2$ elements must be equal to the following value.
\[ W_A + W_B = \frac{n(n+1)}{2} \]  

where \( n = n_1 + n_2 \). Equation (15) can be used for checking if the ranking procedure has been correctly performed.

When each of the samples has less than 10 elements \((n_1 \text{ and } n_2 \leq 10)\), then the determination of significance levels can be found in [10], i.e., the lower and the upper critical value of \( W \). In two-way tests (when \( M_1 = M_2 \) is tested), if the calculated value of the test statistics \( W_A \) is lower than, or equal to the lower critical value, or if it is higher than, or equal to the upper critical value, \( M_1 \) and \( M_2 \) differ, i.e., these two samples do not belong to the same basic set, with a predefined significance level \( \alpha \).

In a one-way test (when \( M_1 < M_2 \) is tested) the critical region lies to the left, so \( H_0 \) is rejected if the obtained value of \( W_A \) is lower than or equal to the lower significance level. A rank sum in the first sample that is too small indicates that the sample originates from a set with a lower median value. Inversely, when a one-way test with the alternative \( H_1: M_1 > M_2 \) is used, the critical region lies to the right, and the null hypothesis is rejected if the realized value of the test statistics \( W_A \) is larger than or equal to the upper significance level.

**Wilcoxon rank-sum test for large samples**

The \( W_A \) test statistics for large samples has an approximately normal distribution, with the arithmetic mean of \( n_1(n + 1)/2 \) and variance of \( n_1n_2(n + 1)/12 \). Hence, for sample sizes of \( n_1 \) and \( n_2 > 10 \), the null hypothesis is tested by an approximate test statistics \( Z \), calculated as

\[ Z = \frac{W_A - \frac{n(n+1)}{2}}{\sqrt{\frac{n_1n_2(n+1)}{12}}} \]  

Bounds \((-z, +z)\) that need to contain the approximate test statistics \( Z \) are then determined. The value of the standard normal (Gaussian) distribution in \( z \) is first determined from

\[ P(-z < Z < z) = 2F(z) - 1 \]  

where \( P(-z < Z < z) \) is the probability of \( Z \) being within the range \((-z, +z)\), while \( F(z) \) is the value of the distribution function in \( z \). Using tabulated values of \( F(z) \) for the standard normal distribution, the upper bound \( z \) is determined, and then, symmetrically, also the lower bound \(-z\).

In a two-way test (when \( M_1 = M_2 \) is tested), if the calculated value of the approximate test statistics \( Z \) is lower than or equal to the lower bound \(-z\), or if it is larger than or equal to the upper bound \( z \), \( M_1 \) and \( M_2 \) differ, i.e., these two samples don’t belong to the same basic set with a significance level \( \alpha \) determined from

\[ P(-z < Z < z) = 1 - \alpha \]  

In a one-way test (when \( M_1 < M_2 \) is tested) the critical range lies to the left, so the samples don’t belong to the same basic set if the obtained value of \( Z \) is less than or equal to the lower bound \(-z\). Inversely, when a one-way test with the alternative \( H_1: M_1 > M_2 \) is used, the critical region lies to the right, and the samples do not belong to the same basic set if the realized value of the approximate test statistics \( Z \) is larger than or equal to the upper bound \( z \).

**VERIFICATION OF THE APPLICABILITY OF THE STATISTICAL METHODS TO NUCLEAR COUNTING RESULTS**

In order to verify the efficiency of the previously described statistical methods concerning nuclear counting, a statistical analysis of background radiation was performed, expected to follow the Poisson distribution, as well as the analysis of nuclear counting of \(^{137}\)Cs radioactive decay, expected to follow the normal distribution. The “number of pulses” random variable was determined from the readings of a Geiger-Müller counter with no anticoincidence protection. Background measurements were performed in 10 s intervals, providing a total of 1000 values for the “number of pulses” random variable.

The statistical sample obtained is, thus, presented in the form of an experimental probability density function (PDF), obtained by normalizing the histogram of measured values, along with the PDF of the corresponding theoretical distribution with parameters determined from the measurement results. Assuming that the hypothesis on the sample’s distribution is true, these two curves should coincide [10-12].

The sample of the “number of pulses” random variable was treated both with and without the application of Chauvenet’s criterion and of the \( U \)-test. The \( U \)-test was applied by dividing 1000 statistical samples into 20 consecutive sub-samples with 50 values of the random variable each, which were then tested to determine if they belong to the same random variable as the first sample. Statistical sub-samples that did not pass the test (at the 5% significance level) were rejected.

The obtained results are shown in figs. 3 and 4.

Assuming a chosen theoretical distribution is adequate for the investigated random variable, the normalized and the theoretical curves should coincide. As figs. 3 and 4 clearly demonstrate, the coincidence is much better when the proposed procedure of rejecting spurious results is applied to the samples. This proves the point that the application of the proposed additional statistical methods excludes random fluctuations of the radioactive source or those of the background radiation from the total statistical sample, as well as possible inadequacies in the experimental setup.
In this paper, a method of purifying the results of nuclear counting is proposed. The proposed methods of applying Chauvenet’s criterion and the $U$-test enable us to obtain a statistical sample of nuclear counting without suspicious single results or without suspicious groups of consecutive results, as it often happens in practice. For, in practice, in nuclear radiation detection, it often happens that inside a single interval of measurement significant fluctuations appear because of a significant shower of secondary cosmic rays or, due to greater fluctuations in the value of the counter’s dead time. In practice, the fluctuation inside one group of consecutive counting results appears due to a certain systematic error of the experiment in which, over a longer detection period, some other source of radiation is superimposed to measured data. These two effects corrupt the statistical regularity of the detection and the statistical sample. The proposed methods for the treatment of statistical samples have proven to be extremely effective and have yielded a high agreement between the theoretical distribution of random variables and the corresponding experimentally obtained random variable.

ACKNOWLEDGEMENT

Under contract 171007, this work has been supported by the Ministry of Education and Science of the Republic of Serbia.

REFERENCES

Ц. В. Долићанин, Ђ. Станковић, Дина Ђ. Долићанин, Борис Б. Лончар

Статистичка обрада резултата нуклеарног бројања

Пошто тачан тренутак када језгро подлеже радиоактивном распаду не може бити одређен, као и предвиђена због неодређености радиоактивног распада, као и друге случајне ефекти у експерименту, статистички закони играју важну улогу у свим случајевима експерименталне нуклеарне физике. У овом раду описан је метод статистичке обраде резултата нуклеарног бројања добијених експериментално, узимајући у обзир случајне промене које се однесу на врло честе и ретке појаве. При обради мерних података препоручује се најпре одбацити сумњиве случајне промене које извршени критеријумом и тестовима, а потврдиће ли резултати у статистичком узорку прате јединствену статистичку расподелу, утицајући на оба похвала и уточнено на тестовање и анализу резултата. Резултати показвали су да су приложени статистички методи ефикасни и добијени резултати потврђивани статистичким тестовима, уместо чистог емпиричног метода. Резултати потврђују ефикасност експерименталних поставки и потврђују коректност експерименталних узора, као и могуће независности експерименталних поставки и потврђују ефикасност експерименталних узора.

Кључне речи: статистичка обрада, нуклеарно бројање, радиациона флукутација, Y-тест

Received on June 8, 2011
Accepted on July 1, 2011