# Radioactivity Registered With a Small Number of Events

*Victor* Zlokazov[1],[⋆] and *Vladimir* Utyonkov[2],[⋆⋆]

[1] *Joint Institute for Nuclear Research Laboratory of Information Technologies*
[2] *Joint Institute for Nuclear Research Laboratory of Nuclear Reactions*

**Abstract.** The synthesis of superheavy elements asks for the analysis of low statistics experimental data presumably obeying an unknown exponential distribution and to take the decision whether they originate from one source or have admixtures. Here we analyze predictions following from non-parametrical methods, employing only such fundamental sample properties as the sample mean, the median and the mode.

The experiments on the synthesis of the superheavy elements yield chains of the sequential decays of the elements $E_i$, following a scheme: $E_i \rightarrow E_{i+1} \rightarrow E_{i+2} \rightarrow E_{i+3}$ etc. Each examined chain consists of registered decay times $t_i$ of the corresponding element although some chains can be incomplete. These experiments can be performed by different groups using somewhat different methods of producing the nuclei under study, registration, and analysis of the obtained data. The goal of the present analysis is to estimate the statistical behavior of the detected elements and to test the results of different groups for compatibility.

First, we have to answer the following question: should we analyze each individual component in the chains or analyze the chain as a whole? To make it more comprehensive, let us recall the basic properties of the radioactive decays [1].

- Any $k$ decays belong to the same type, are independent, occur in non-intersecting time points, and their probability within any time interval $[t_0, t_0 + \Delta t]$ does not depend on the choice of $t_0$;

- The decays are labeled as 'rare events', i.e., the probability of 2 and more events for any small $\Delta t$ is infinitesimal compared with the probability of 0 or 1 events;

- Markov property of decays – the probability of observing any number of decays after any moment $t$ does not depend on the origin of the element (it means on its prehistory).

Thus, our problem is: being given the experimental events $t_{iEc}$, where $i$ is the number of event, $E$ is the name of the isotope, $c$ is the number of team performing the experiment,

- estimate the parameters of each group of data $t_{iEc}$, $i = 1, \ldots, n_{Ec}$ for each couple $Ec$

- test these estimates $\hat{t}_{Ec}$ for each $E$ and $c$ for compatibility.

However, do all the methods follow the above rules? For instance, if the third condition (Markov property) is not satisfied, we should not analyze separate events, but consider the chain as a whole (for each $E$ and $c$), which may complicate the analysis. Therefore, we must first test the definitely uncorrelated data $t_{iEc}$ of $E_j$ and the corresponding $t_{iEc}$ of $E_{j+1}$.

⋆e-mail: zlokazov@jinr.ru
⋆⋆e-mail: utyonkov@jinr.ru

Given two simulated samples of uncorrelated data $A_1 = (t_{11}, t_{12}, \ldots, t_{1n})$ and $A_2 = (t_{21}, t_{22}, \ldots, t_{2n})$ with the same exponential distribution function, their sample correlation coefficient $r$ is

$$r = \sum_{i=1}^{n} (t_{1i} - m_1)(t_{2i} - m_2)/(\sigma_1 \cdot \sigma_2)/n, \tag{1}$$

where $m_1, m_2, \sigma_1, \sigma_2$ are expectations and sigmas (standard deviations) of the both distributions.

In an ideal case (infinite $n$, absolutely exact $m_1, m_2, \sigma_1, \sigma_2$) we should have $r = 0$. However, in our case of finite data, the calculations of $r$ will result in a nonvanishing value. The Monte-Carlo tests showed that the main factor defining the spread of the of sample values of $r$ is the size $n$. Table 1 collects the depending on $n$ minimum, maximum, mean and unspecified values of $r$, respectively.

| $n$ | Min | Max | Mean | MAbs |
|---|---|---|---|---|
| 10 | −0.1335 | 0.3646 | 0.0865 | 0.1132 |
| 20 | −0.3781 | 0.2673 | −0.0200 | 0.1024 |
| 30 | −0.1011 | 0.1312 | 0.0054 | 0.0367 |
| 50 | −0.1043 | 0.2531 | 0.0091 | 0.0489 |
| 80 | −0.0661 | 0.1659 | 0.0184 | 0.0375 |
| 100 | −0.0623 | 0.1895 | 0.0269 | 0.0394 |

The table demonstrates occurrence of significant deviations from zero of the correlation coefficients for finite samples (especially at values $n = 10$ and 20).

A rigorous mathematical study of the distribution of the sample correlation coefficients can be found in [2], but it covers only asymptotics of normal event distributions and is described by a very cumbersome formula, including infinite series, that can be hardly used in practice.

However, based on results of Table 1, we can specify an interval (e.g. $[-0.4, 0.4]$) and say that if the obtained sample correlation coefficient does not exceed this range, the tested sample does not contradict the hypothesis that the data is uncorrelated.

To verify if the following radioactive decays are correlated or not, we can use the data with rather good statistics, e.g. $^{288}$Mc $\rightarrow$ $^{284}$Nh $\rightarrow$ $^{280}$Rg; 63 events for each decay step were selected from [3, 4]. The performed analysis gave the estimate $r = -0.18$. So the real practice shows examples that do not reject the Markov property of the successive decays.

As for the data discussed in [5] (11 events in each sample) for the transition 'element 115 $\rightarrow$ element 113', we obtained $r = -0.11$ and for the samples 'element 113 $\rightarrow$ element 111', $r = +0.30$. Therefore, we can accept the hypothesis that the successive decays are uncorrelated and we can use for our analysis separate events in each data group rather than the fixed chains of events.

Furthermore, we can estimate the parameters $T$ of all the exponential probability distributions of decay times $t_i$:

$$P(t_i < t) = P(t, T) = 1 - \exp(-t/T), \tag{2}$$

where the expectation of $t_i$ is $T$ and its variance is $T^2$.

The sum of the times of $m$ decays $S = \sum_{i=1}^{m} t_i$ has the real $(m, T)$-gamma distribution [2]. Its density function is

$$g(t, m, T) = \begin{cases} \frac{t^{m-1}}{T^m (m-1)!} \exp(-t/T) & \text{for } t \geq 0; \\ 0 & \text{otherwise,} \end{cases} \tag{3}$$

where $m$ is a positive integer, and a $T$ is positive real. Let's consider a quantity $S_m = S/m$. The density of the $S_m$ distribution is $g_m(t) = m \cdot g(mt, m, T)$, and its mean and the variance are equal to $T$ and to $T^2/m$, respectively. The maximum of $g_m(t)$ (let it be $t_x$) is reached at the root of the equation

$$m \left( \frac{(m-1)(mt)^{m-2}}{T^m (m-1)!} - \frac{(mt)^{m-1}}{T^{m+1}(m-1)!} \right) \cdot \exp(-mt/T) = 0,$$

from which we obtain $t_x = (m-1)T/m$.

If $m$ is small, this maximum can be rather far from the mean $T$. Therefore, the gamma-distribution is not favorable to the low statistics – it is the case where 'the most probable is not the most expected'. If $m$ tends to the infinity, $g_m(mt, m, T)$ tends slowly to the normal density. The statistic $S_m$ is sufficient for (2) [5, 6] and can serve as an estimate of the parameter $T$. At the same time it is the maximum likelihood estimate of $T$ and it is efficient, i. e., having the variance asymptotic $1/n$.

For the two sets of decays $A_1 = (t_{11}, t_{12}, \ldots, t_{1m})$, $A_2 = (t_{21}, t_{22}, \ldots, t_{2m})$, two hypotheses are tested

1. $H_1 : A_1$ and $A_2$ have the same distribution;

2. $H_2 : A_1$ and $A_2$ have different distributions.

We can use a conventional approach to solve this problem by building the confidence intervals for the estimated means $T$ (Student's test) and for the variances $\sigma^2$ ($\chi^2$ test):

$$\hat{T} - t_{\alpha/2,m-1} * \hat{s}/\sqrt{m} < T < \hat{T} + t_{\alpha/2,m-1} * \hat{s}/\sqrt{m},$$

$$(m-1)\hat{s}^2/\chi^2_{\alpha/2,m-1} < \sigma^2 < ((m-1)(\hat{s}^2/\chi^2_{1-\alpha/2,m-1},$$

where $\hat{T} = \sum_{i=1}^m t_i/m$, $\hat{s} = \sqrt{\sum_{i=1}^m (\hat{T} - t_i)^2/(m-1)}$, $t_{\alpha,m-1}$ and $\chi^2_{\alpha,m-1}$ are the values of the Student's and $\chi^2$ standard random quantity, that corresponds to the $\alpha$ probability and the number of degrees of freedom $m-1$.

It is reasonable enough to accept $H_1$ if the confidence intervals $I_1$ and $I_2$ overlap and the probability of the interval overlapping is sufficiently large. Otherwise, we can calculate the so called Type I and Type II errors for rejecting a true hypothesis and accepting the wrong one, and, after analyzing them, make the corresponding decision.

We shall now consider the concept of the optimal confidence interval $[a, b]$ (OCI) for testing the hypotheses [7]. Such an OCI should have the maximum accuracy of a testing criterion which means: the minimal difference $b - a$ and, at the same time, the probability of the events to belong to the interval $[a, b]$ should be maximum. Since these conditions contradict each other, an OCI should be determined from a compromise between two requirements [7]:

- for a given length $b - a$ to find an interval with the best ratio 'pro/contra';

- for a given ratio 'pro/contra' to find an interval of the shortest length $b - a$.

Very often one takes the interval $[T - \sigma, T + \sigma]$, the most popular for the normal distribution. For the gamma distribution it will be $[T - T/\sqrt{n}, T + T/\sqrt{n}]$; however, the convergence of the gamma distribution to the normal one is slow so that this interval will be asymmetric even at large $n$.

Apart from these criteria, we can build an OCI from the requirement that the physical meaning of the interval $[a, b]$ and its bounds $a$ and $b$ should be the most obvious and natural.

One of the ways to build such an OCI is to make use of the so called order statistics, which for the distribution (3) can be derived as easily integrable analytical functions [7]. Following [2] we define the following order statistics. Denote the minimum value in the sample $S$ as $u_1$; and denote the maximum value in the sample $S$ as $u_m$; their expectations $\hat{u}_1 = \hat{E}u_1$ and $\hat{u}_m = \hat{E}u_m$ are the order statistics of the 1st and the last rang, the meanings of which are:

1. $\hat{u}_1$ is the average minimum value in the sample;

2. $\hat{u}_m$ is the average maximum value in the sample. They can be used as the most 'natural' bounds of the average decay time in the sample.

An additional means to test the two distributions for congruence is to take the expectation $E_d$ and the variance (sigma) $\sigma_d$ of the normalized differences of both sets $t_{1i}$ and $t_{2i}$ and build the confidence interval $[-\sigma, \sigma]$ and see whether the $E_d$ falls into it. Unfortunately, the formula for the variance of

the difference of the two gamma-distributed random quantities has no simple closed form, so that it is difficult to estimate the trustworthiness of this test.

Now, we describe a method to solve a very important problem: is the analyzed data really a sample from a single exponential distribution or from a mixture?

This method is based on checking the sample relations between such fundamental characteristics of any event distribution as the mode, the median and the mean. In the case of exponentials these relations are uniquely defined and are different for a single component and a mixture of components.

The relation between the mean $E$ and the median $M$ for a single exponential is $M = E \cdot \ln 2$. For the ratio $K$ of 'median/mean' in finite samples (e.g. $n = 15$) the statistical test gives $K = 0.7 \pm 0.17$ for a confidence interval of 67 %. However, if the data consist of the decays of two (or more) sources, $K$ can differ from $\ln 2$ by several orders. Therefore, $K$ is both a simple and a rather reliable indicator of whether the data originate from a single source or from a mixture of components.

## Discussion and Conclusions

The literature covering the methods for the problem solution is very large (e.g. one can note the excellent paper [8]), but usually they are oriented to the analysis of large data volume size about which there is a certain amount of a priori information available. Such are all the methods based on the use of the regression techniques or likelihood relations. To advance in the data analysis with low statistics the accent was made on the non-parametric methods, employing only such fundamental sample properties as the sample mean, the median and the mode.

The report confirms that a radioactive process really possesses at least the weak form of the Markov property - the uncorrelatedness of the decays that is the necessary condition for a correct use of the mathematical techniques. It provides an analysis of the various methods of constructing of optimal and easily interpretable confidence intervals, using the order statistics and suggests a simple non-parametric criterion for testing the exponential data for admixtures by comparing the sample median with the sample mean. As example, this criterion applied to the data [9] reveals that some of them are most probably not pure.

## Acknowledgments

## References

[1] B. V. Gnedenko, *Theory of Probability* (English translation, Mir Publishers, 1978)

[2] J. E. Freund, *Mathematical Statistics* (Prentice-Hall, Inc. Englewood Cliffs, N. J., 1962)

[3] Yu. Ts. Oganessian et al., Phys. Rev. C 87, 014302 (2013); D. Rudolph et al., Spectroscopy of Element 115 Decay Chains, Phys. Rev. Lett. **111**, 112502 (2013)

[4] J. M. Gates et al., Phys. Rev. C **92**, 021301(R) (2015)

[5] U. Forsberg et al., Nucl. Phys. A **953**, 117 (2016)

[6] S. S. Wilks, *Mathematical Statistics* (Princeton, N.J., Princeton University Press, 1947)

[7] V. B. Zlokazov, Physics of Particles and Nuclei Letters **12**, 2, 262–268 (2015)

[8] M. Stehlik and H. Wagner, IFAS Research Paper Series 2009-39 (January 2009)

[9] V. B. Zlokazov and V.K. Utyonkov, J. Phys. G: Nucl. Part. Phys. 44 (2017) 075107 (13pp)