Sándor Nagy

STOCHASTICS AND NUCLEAR MEASUREMENTS

(Summary of Probability Theory with Nuclear Applications)

For DA who had a Heart of Gold

LIBRARY OF NAGY'S E-BOOKS
ELTE, KI, Budapest, 2010

CONTENTS

Introduction

Louis Lyons writes in the preface of his book 'Statistics for Nuclear and Particle Physicists' (Lyons 1986) that 'one cannot learn statistics simply by reading a book on the subject'. This applies even more so to a wider field as we use the term 'stochastics'[1] here. Therefore we must be content with summarizing the relevant facts of probability and statistics and pointing out possible connections with nuclear measurements. As regards sampling distributions, parameter estimation etc. we refer to the above monograph as well as to Press *et al.* 1999. We will only make an exception with the χ^2 distribution, which is a valuable goodness-of-fit tool in nuclear spectroscopy. We will also cite a few examples of the nuclear applications of stochastic processes, a field usually excluded from short studies of this type.

Nuclear aspects (as well as some other important points of reference) will be phrased in the forms of **remarks** numbered like this: **(#1)**, **(#2)**, etc. Such remarks also serve occasionally as interface between the (mathematical) terminology used in this text and the terminologies used by other authors coming from different fields of nuclear science. The numbering of the remarks may also help to follow a train of thought that is unfolded over several chapters and sections.

1. Distributions of Random Variables

To establish uniform notation and to provide the reader with a convenient reference/vocabulary we will cite some of the concepts and formulae of probability theory and statistics that we will use later.

1.1. Measures of the 'location' of a distribution

Expected value. Among the **location parameters**, designed to show where the 'bulk' of a distribution is concentrated, the expected value (or, as we will alternatively call it, the **mean**) is considered as standard (provided that it exists). We will usually denote it by μ. If we want to make it clear that we are talking about the expected value of the **random variable** X, then we will use the notation $E(X)$. Further notations used in the same sense are $\langle X \rangle$ and μ_X.

(#1) Frequently used synonyms for the expected value/mean are *expectation value*, *mathematical expectation* or just *expectation*. In physics and other fields of science (except mathematics), the expected value is often referred to, in a rather careless way, as 'average', but we will reserve this term for the estimate of the mean as defined by Eq. (29). In physics, it is also customary to use the symbol $\overline{X}$ for the mean, but we will reserve the horizontal overbar to denote the sample mean, in other words, the 'real' average.

The expected value of a **discrete distribution** is calculated from its **mass function** $p(x) \equiv P(X = x)$, where P means probability. The formula is:

$$E(X) \equiv \sum_{\forall i} x_i p_i \tag{1}$$

[1] The term 'stochastics' in the title as we mean it roughly translates to 'random features'. So it refers to anything related to probability theory, statistics and, of course, stochastic processes. Hans-Otto Georgii defines 'stochastics' as 'a term comprising both probability theory and statistics' in Footnote 1 of his book titled 'Stochastics: Introduction to Probability and Statistics' (Georgii 2008).

where p_i is the weight of the ith **spectrum point**[2] x_i. For spectrum points $P(X = x_i) \equiv p_i$. If x is not a spectrum point, then $P(X = x) = 0$.

The **distribution function** of discrete distributions is calculated from the following sum:

$$F(x) \equiv \sum_{\forall i: x_i < x} p_i \tag{2}$$

For **normalized distributions** that we are dealing with: $F(\infty) = 1$.

Integral valued random variables are an important class of discrete distributions. Their spectra consist of the integers: 0, 1, 2, … The expected value is:

$$E(X) \equiv \sum_{i=0}^{\infty} i p_i \,. \tag{3}$$

For **continuous distributions** the expected value of X is calculated from the **density function** f as follows:

$$E(X) \equiv \int_{-\infty}^{+\infty} x f(x) \mathrm{d}x \,. \tag{4}$$

The **distribution function** of continuous distributions is calculated from the following integral:

$$F(x) \equiv \int_{-\infty}^{x} f(u) \mathrm{d}u \,. \tag{5}$$

For **normalized distributions** (that we are dealing with) again we have: $F(\infty) = 1$.

It follows from Eq. (5) that the density function is the derivative of the distribution function:

$$f(x) = \frac{\mathrm{d}\, F(x)}{\mathrm{d}\, x} \,. \tag{6}$$

(#2) Note that in various fields of science, when people talk about **averaging a physical quantity** what they usually mean is calculating the expected value of that quantity (considered but not necessarily declared as a random variable). The truth is revealed by the fact that they use formulae like Eqs. (1) and (4) for those calculations. The mass function and the density function are often not distinguished either, but they are referred to by the same expression like *probability density* or *differential distribution function*[3] and sometimes, rather loosely, 'distribution function'[4] or just 'distribution'[5]. The 'distributions' they do their 'averaging'

[2] The expression 'spectrum point' refers to any of the 'allowed' values of a discrete random variable in this context, and therefore it has nothing to do with the nuclear spectra discussed in Chapter 5.

[3] The explanation of the adjective 'differential' is made clear by Eq. (6).

[4] If the density function or the mass function is referred to as 'distribution function', then the 'real' distribution function is normally called *integral distribution function*. The reason for the name is clear from Eq. (5).

[5] As a matter of fact, very often, no reference is made to any distribution at all, although careful analysis of the problem reveals that some of the quantities are *un-normalized density functions*. As an example we mention the various quantities—actually different types of joint density functions of multivariate distributions—all referred to by the same term *neutron flux* and denoted by the same symbol (either φ, ϕ or Φ) no matter how many and which of the possible variables (space coordinates, solid angle, speed or energy or, alternatively, lethargy) are considered or made disappear by integration. To make things even more confusing, the only related quantity 'defined' by IUPAP bears the name *neutron flux density*, (dimension: number of neutrons per square centimeter per second), although this is the least 'density-function-like' of the whole family of related quantities. Apropos of

with are in fact **un-normalized density functions** most of the time. That is, in the continuous case, e.g., the 'averaging' in physics etc. typically goes like this:

$$\langle x \rangle = \frac{\int_{-\infty}^{+\infty} x g(x)\,\mathrm{d}x}{\int_{-\infty}^{+\infty} g(x)\,\mathrm{d}x} = \frac{1}{A}\int_{-\infty}^{+\infty} x g(x)\,\mathrm{d}x\,. \tag{7}$$

where A^{-1} is called the **normalizing factor** converting the un-normalized density function $g(x)$ to the normalized one, $f(x) = A^{-1}\, g(x)$, for which:

$$\int_{-\infty}^{+\infty} f(x)\,\mathrm{d}x = \frac{1}{A}\int_{-\infty}^{+\infty} g(x)\,\mathrm{d}x = \frac{A}{A} = 1\,. \tag{8}$$

as it should be. (Needless to say that the integration limits can be different from those indicated.)

(#3) Note also that Eq. (7) is written in the way as is usually done outside mathematics, i.e., the symbol x is used both for denoting the random variable (on the left) and for indicating the assumed values of that same variable (on the right). (According to the convention we try to follow in this text, the notation should be X on the left.) However, this does not affect the results of such calculations at all, only the notation is somewhat confusing (to a mathematician anyway), but that is how most people do it in practice.[6]

(#4) Another type of 'averaging' to be mentioned here is actually beyond the scope of this text. In *quantum physics*, the most common symbolism for the calculation of the expected value of a physical quantity q (characterized by the operator $\boldsymbol{q}$) in the quantum state described by the (complex) wave function $\Psi(\boldsymbol{r})$ is this:

$$\langle q \rangle = \langle \Psi | \boldsymbol{q} | \Psi \rangle \equiv \int \Psi^* \boldsymbol{q}\, \Psi\, \mathrm{d}V\,. \tag{9}$$

where the integration goes by the volume element $\mathrm{d}V$ (over the whole space) and Ψ^* is the complex conjugate of Ψ. Here the product $\Psi^*\Psi$ is referred to as **probability density** too. It is indeed normalized to 1 as one would rightly expect from a density function:

$$\int \Psi^*\Psi\, \mathrm{d}V = 1\,. \tag{10}$$

Median. The median $X_{1/2}$ is the point along the x axis which represents the middle of the distribution in the sense that X values smaller and greater than $X_{1/2}$ are equally probable. For a continuous distribution this is the point where the distribution function $F(x)$ reaches half of the maximum:

$$F(X_{1/2}) = \frac{1}{2}\,. \tag{11}$$

(#5) As a familiar example of the median in nuclear science we should mention the half-life $T_{1/2}$ of a radionuclide. (See Subsection 3.4.1 as well as the right-hand panels in FIGURE 1.)

---

dimension: when in doubt as regards what type of neutron flux we are encountered with, we propose *dimensional analysis* as a guide.

[6] As a matter of fact, the author himself has to concentrate very hard not to fall out of his chosen role as a 'mathematician', but he probably fails occasionally anyway.

For **statistical samples** the median is either the middle element (for an odd sample size) or the average of the two middle elements (for an even sample size) when the elements are arranged in increasing order (ordered sample).

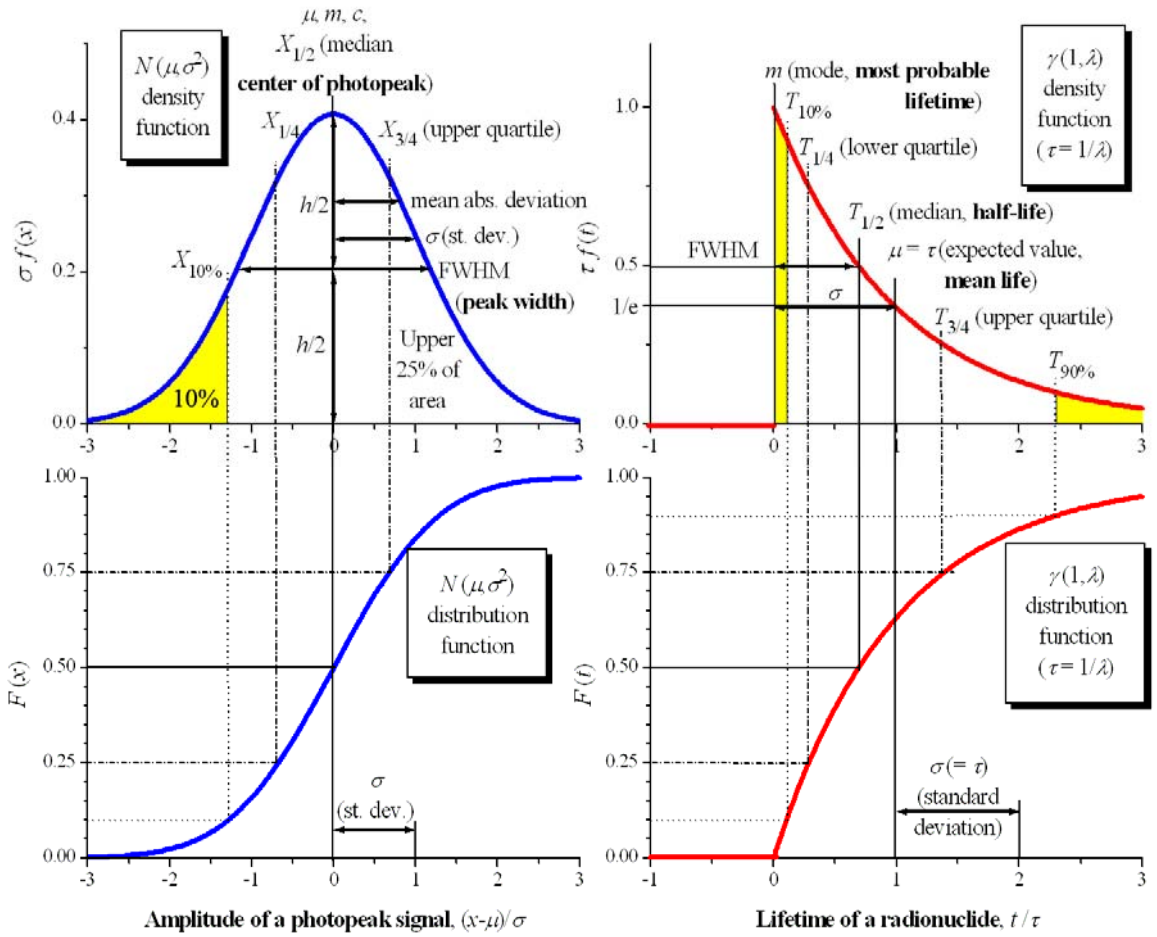

FIGURE 1. Presentation of the location and dispersion parameters of a symmetric (normal) and an asymmetric (exponential) distribution by the help of the density function (upper panels) and the distribution function (lower panels). The shaded areas under the density curves represent 10% of the total area under them in accordance with the meaning of the 10 percentile ($X_{10\%}$) and with the relationship between the density function and the distribution function shown by Eq. (5). Note that for an asymmetric distribution the '+' and '–' error intervals measured from the mean to complementary values of quantiles/percentiles are different, whereas for a symmetric distribution the estimate of the mean obtained from a single observation of X can always be given as $x \pm \Delta x$ no matter how the error Δx is calculated (e.g. it can be given in terms of the 95% (+) and 5% (–) percentiles, or it can be set equal to 2σ as usual). In practice, most people give their measured or calculated 'results' with the '±' convention without paying any attention to the actual shape of the 'error distribution'.

Mode. The mode m represents the local maximum (or one of the local maxima) of the density/mass function. It is usually referred to as the **most probable value** in physics and chemistry. (In the case of discrete distributions only the spectrum points are considered when looking for maxima.) If there is only one maximum, the distribution is called *unimodal*, if there are two maxima, the distribution is called *bimodal* etc.

Connection between location parameters. The existence of expected value depends on the convergence of the sum/integral from which it is calculated. If the expected value does not exist, the rest of the location parameters can still be used as substitutes. (For multimodal

distributions the modes can be actually more useful than the single expected value because they also convey information about the shape of the distribution.)

For **symmetrical unimodal distributions** the location parameters coincide with the **center of symmetry** c of the mass/density function:

$$\mu = X_{1/2} = m = c\,. \tag{12}$$

For **asymmetrical unimodal distributions** (see the upper right graph in FIGURE 1) the following 'rule of thumb' is often cited: The median divides the distance between the mode and the mean in the approximate ratio 2:1, i.e.:

$$(X_{1/2} - m) : (\mu - X_{1/2}) \approx 2:1\,. \tag{13}$$

1.2. Useful tools: generating function and characteristic function

Law of the Unconscious Statistician (Goodman 1988). The expected value of the random variable Y obtained by the function/transformation $Y = g(X)$ is calculated analogously to Eqs. (1) and (4) using the mass/density function of X:

$$E(g(X)) = \sum_{\forall i} g(x_i)p_i\,; \qquad E(g(X)) = \int_{-\infty}^{+\infty} g(x)f(x)\mathrm{d}x\,. \tag{14}$$

Special applications of the above formula are the **momenta** of a distribution. The ***n*th moment** is defined as follows:

$$M_n \equiv E(X^n)\,. \tag{15}$$

In particular, the **first moment** is equal to the expected value:

$$M_1 \equiv E(X) = \mu\,. \tag{16}$$

The square root of the second moment (also called mean-square)—one of the 'effective means' used—is usually referred to as the **root mean square** or **rms**:

$$\mathrm{rms} \equiv \sqrt{M_2} = \sqrt{E(X^2)}\,. \tag{17}$$

Beside the rms, other types of **effective means** can be defined by the formula

$$\text{effective mean} \equiv g^{-1}\{E[g(X)]\} \tag{18}$$

where g^{-1} is the inverse of the function g. As a further example of the effective means we mention the **harmonic mean**, i.e. the reciprocal of the mean-reciprocal:

$$\text{harmonic mean} \equiv [E(X^{-1})]^{-1}\,. \tag{19}$$

(#6) Note that the **reduced mass** $m_\mathrm{r} = m_1m_2/(m_1+m_2)$ of two masses m_1 and m_2 often cited in this text is exactly one half of their harmonic mean as the reader can easily check by assigning the same weight $p_i = 1/2$ when applying Eq. (1) to the simple calculation. Division by two causes the reduced mass to deserve its name inasmuch as it is always less than either of the masses. On the other hand, the harmonic mean itself always lies between m_1 and m_2 as one can rightly expect from a mean of any type. (See also Section 5.4.)

Generating function. The generating function is defined for integral valued random variables the spectra of which consist of 0, 1, 2, … with weights p_0, p_1, p_2, … The generating function of X is defined as the expected value of the exponential function s^X:

$$G(s) \equiv E(s^X) = \sum_{k=0}^{\infty} p_k s^k \ . \tag{20}$$

Since the distribution is normalized, $G(1) = 1$. The generating function (just like the mass function) contains all information about the distribution. For instance, the expected value and the variance of X (see later on) can be calculated from its derivatives as follows:

$$E(X) = \mu = G'(1)\,; \qquad D^2(X) = G''(1) + \mu(1-\mu)\,. \tag{21}$$

Characteristic function. The characteristic function of continuous distributions is the expected value of the complex exponential function $\mathrm{e}^{\mathrm{i}uX}$ (where $\mathrm{i}^2 = -1$). In other words it is the **Fourier transform** of the density function f:

$$\varphi(u) \equiv E\left(\mathrm{e}^{\mathrm{i}uX}\right) = \int_{-\infty}^{+\infty} \mathrm{e}^{\mathrm{i}ux} f(x)\,\mathrm{d}x\,. \tag{22}$$

Since the distribution is normalized, $\varphi(0) = 1$. The characteristic function (just like the density function) contains all information about the distribution. For instance, the **expected value** and the **variance** of X can be calculated from its derivatives as follows:

$$E(X) = \mu = \frac{1}{\mathrm{i}}\varphi'(0)\,; \qquad D^2(X) = -\varphi''(0) - \mu^2\,. \tag{23}$$

Both the characteristic function and the generating function represent a transformation T for which the following theorem holds (Korn and Korn 1968). Let X and Y denote independent random variables with density/mass functions f_X and f_Y. Let f_{X+Y} denote the density/mass function of the sum $X+Y$. Then we have:

$$T(f_{X+Y}) = T(f_X)\,T(f_Y)\,. \tag{24}$$

The above formula serves as a basis for the **deconvolution** as well as for the **addition theorems** that we will cite later on in connection with different distributions.

1.3. Measures of the ‘dispersion’ of a distribution

Variance. Among the **dispersion parameters** characterizing the ‘spread’ of a distribution, the variance is considered as standard (provided that it exists). The variance is denoted by σ^2 or, if we wish to emphasize that we are talking about the variance of the random variable X, the notation is $D^2(X)$, $Var(X)$, or σ^2_X. The variance of X is calculated as follows:

$$D^2(X) \equiv E[(X-\mu)^2] = M_2 - \mu^2\,. \tag{25}$$

where $M_2 \equiv E(X^2)$ is the **second moment** of X. (The **first moment** is $M_1 \equiv \mu$.) **Standard deviation.** The standard deviation of X is calculated from the variance:

$$D(X) \equiv \sqrt{D^2(X)}\,. \tag{26}$$

Further notations: σ or σ_X. The **relative deviation** is the ratio of the standard deviation to the expected value:

$$\sigma_{\text{rel}} \equiv \frac{\sigma}{\mu}. \tag{27}$$

Mean absolute deviation. The (mean) absolute deviation of X is:

$$\text{Abs D}(X) \equiv E\big(|X - \mu|\big). \tag{28}$$

Whenever the expected value exists, so does the absolute deviation. (The standard deviation may not exist, even if the expected value exists.)

Interquantile range. This type of 'measure' works even if the expected value does not exist.

The ***p* quantile/fractile** is a (not necessarily unique) point X_p along the x-axis, where the distribution function reaches a given p fraction of the maximum, i.e. where $F(X_p) = p$. Special types of quantiles are the **quartiles** (p = 1/4, 1/2, 3/4), the **deciles** (p = 0.1, 0.2, … 0.9) and the **percentiles** (p = 0.01, 0.02, … 0.99).

Quantiles can be used for the characterization of the spread of the distribution by giving the distance between the points X_p and X_{1-p}. For instance, the **10-90-percentile range** means the difference $X_{0.9}-X_{0.1}$, where $X_{0.1}$ and $X_{0.9}$ are the 10 and the 90 percentile, respectively. It follows from the definition that about 80% of the observed values of X are supposed to lie in this range and about 20% outside.

The **interquartile range** characterizes the spread of the distribution with the distance between the lower and the upper quartiles: $X_{3/4}-X_{1/4}$. (The lower, middle and the upper quartiles mean the 25, 50 and 75 percentiles, respectively.)

According to the above terminology, the median can be either considered as the middle quartile ($X_{2/4}$), or as the 50 percentile ($X_{0.5}$) of the distribution.

Halfwidth. The spread of a unimodal (continuous) distribution is sometimes characterized by its halfwidth meaning its **full width at half maximum** abbreviated as **FWHM**. The term refers to the total width of the density function (peak) between the points, where its height is half of the maximum.

The **relative width** is the ratio of the FWHM to the expected value: $\text{FWHM}_{\text{rel}} = \text{FWHM} / \mu$.

1.4. Estimation of expected value and variance

Estimation of the expected value. The expected value is estimated by the **average** (also called the **sample mean** or **arithmetic mean**):

$$\hat{\mu} = \overline{X} \equiv \frac{1}{n}\sum_{k=1}^{n} X_k \tag{29}$$

where '$\hat{\mu}$' reads: 'the **estimate** of μ'. According to Eq. (45), this estimate is **unbiased**, that is

$$E\big(\overline{X}\big) = \mu. \tag{30}$$

(#7) As mentioned before, in physics, the horizontal overbar above a variable is often used to indicate its expected value, but we try to reserve this notation for the average only.

According to Eq. (47), the **uncertainty of expected-value estimation** is determined by the variance of the average:

$$D^2(\overline{X}) = \frac{1}{n} D^2(X) \equiv \frac{1}{n}\sigma^2; \qquad D(\overline{X}) = \frac{1}{\sqrt{n}} D(X) = \frac{1}{\sqrt{n}}\sigma \tag{31}$$

i.e. if we want to increase the accuracy of expected-value estimation by a factor of 2, 3 or N, then we need to average 4, 9, N^2 times as many data, respectively (see FIGURE 3).

Estimation of variance. The variance is normally estimated by the following formula of the **empirical variance**/**sample variance** (note that the sample mean $\overline{X}$ is determined from the same set of data as the sample variance itself):

$$\hat{\sigma}^2 = s^{*2} \equiv \frac{1}{n-1}\sum_{k=1}^{n}(X_k - \overline{X})^2 . \tag{32}$$

The above formula looks rather surprising having the form of a 'spoiled' average. Its use, however, is justified since the variance estimate thus obtained is **unbiased**, i.e.:

$$E(s^{*2}) = \sigma^2 . \tag{33}$$

If the expected value μ is exactly known, then the 'unspoiled' averaging:

$$\hat{\sigma}^2 = s^2 \equiv \frac{1}{n}\sum_{k=1}^{n}(X_k - \mu)^2 \tag{34}$$

delivers the unbiased estimate for the variance. Note that the estimate calculated from Eq. (32) tends to be greater than that obtained from Eq. (34) by a factor of $n/(n-1)$. Thus the difference is only significant when the sample size n is small. However, for small samples variance estimation is rather meaningless anyway. Therefore in most practical cases the choice between the two formulas is only a matter of taste/convenience (i.e. whichever formula is available it will do). The estimate of standard deviation in either case is:

$$\hat{\sigma} = \sqrt{\hat{\sigma}^2} . \tag{35}$$

Weighted average. It is a common situation that the value of a physical quantity is determined from different types of experiments, and the experimental values (X_k) obtained for those quantities have their own (different) accuracies characterized by the standard deviation (σ_k). When the error distribution of each experimental value can be considered normal, then the **maximum likelihood estimate of the expected value** of the physical quantity is given by the following weighted average (Orear 1987):

$$\hat{\mu} = \frac{\sum_{k=1}^{n} \frac{1}{\sigma_k^2} X_k}{\sum_{k=1}^{n} \frac{1}{\sigma_k^2}} . \tag{36}$$

The variance of the expected-value estimate is now (Leo 1987):

$$D^2(\hat{\mu}) = \frac{1}{\sum_{k=1}^{n} \frac{1}{\sigma_k^2}} . \tag{37}$$

It is easy to verify that if the errors of the X_k variables are equal (i.e. when the experimental values have been determined from the same experiment with the same experimental error $\sigma_k \equiv \sigma$), then Eq. (36) changes over to the usual formula of simple arithmetic average as shown by Eq. (29) whose uncertainty is given by Eq. (31).

1.5. Measures of the 'association' of two distributions

In the following formulas the subscripts 1 and 2 refer to the random variables X and Y, respectively. The **independence** of the random variables X and Y means that their **joint distribution function** is equal to the product of their respective (one-dimensional) distribution functions:

$$F(x,y) = F_1(x)F_2(y)\,. \tag{38}$$

In the case of continuous distributions, independence also means that the **joint density function** is equal to the product of the individual density functions:

$$f(x,y) = f_1(x)f_2(y)\,. \tag{39}$$

The **covariance** of X and Y is defined as:

$$\sigma_{12} \equiv \mathrm{Cov}(X,Y) \equiv E((X-\mu_1)(Y-\mu_2)) = E(XY) - \mu_1\mu_2\,. \tag{40}$$

The **correlation coefficient** of X and Y is given by:

$$\rho_{12} \equiv R(X,Y) \equiv \frac{\mathrm{Cov}(X,Y)}{D(X)\,D(Y)} \equiv \frac{\sigma_{12}}{\sigma_1\sigma_2}\,. \tag{41}$$

The correlation coefficient is a number lying in the interval [−1,+1].
If X and Y are **independent**, then they are also **uncorrelated**, i.e.:

$$\rho_{12} = 0\,. \tag{42}$$

If there is a **linear relationship** between X and Y, then:

$$|\rho_{12}| = 1 \tag{43}$$

i.e. the correlation reaches its maximum. More exactly, if $Y = aX+b$, then:

$$\rho_{12} = \begin{cases} +1 & \text{if} \quad a > 0 \\ -1 & \text{if} \quad a < 0 \end{cases}. \tag{44}$$

Owing to this result, one might conclude (erroneously) that whenever there is a strong relationship between X and Y (e.g. when Y is the function of X), then the variables must be correlated. However the truth is that one can find such distributions and such a strong (however nonlinear) relationship between X and Y, which makes them uncorrelated.

All things considered, we can only state with certainty that **if *X* and *Y* are correlated, they cannot be independent**.

2. Sums and Products, Error Propagation

In this chapter we will rehearse the 'algebra' of random variables from the viewpoint of **error propagation**. **Convolution**, **central limit theorem** as well as **random sums** (i.e. sums of a random number of random variables) are also included here because of their importance in nuclear applications.

2.1. Expected value and variance of linear expressions

Expected value of linear combinations. For any sequence of random variables as well as constants a and b_k we can write:

$$E\left(a+\sum_{k=1}^{n} b_k X_k\right) = a+\sum_{k=1}^{n} b_k \mu_k \,. \tag{45}$$

Corollaries:

Expected value of a constant: ..... $E(a)=a$

Shift in general: ..... $E(a+X)=a+E(X)$

Change of scale in general: ..... $E(bX)=b\,E(X)$

Addition/subtraction in general: . $E(X\pm Y)=E(X)\pm E(Y)$

Variance of the sum of random variables. We can write for any choice of two random variables that:

$$D^2(X\pm Y)=D^2(X)+D^2(Y)\pm 2\mathrm{Cov}(X,Y)\,. \tag{46}$$

For **independent random variables,** the covariances are zero, thus the following formula holds for the **variance of the linear combination** of several variables:

$$D^2\left(a+\sum_{k=1}^{n} b_k X_k\right)=\sum_{k=1}^{n} b_k^2 D^2(X_k)\,. \tag{47}$$

Corollaries:

Standard deviation of a constant: $D(a)=0$

Shift in general: ..... $D(a+X)=D(X)$

Change of scale in general:..... $D(bX)=|b|\,D(X)$

Addition/subtraction *for independence*: $D^2(X\pm Y)=D^2(X)+D^2(Y)$

Note that the **variances add up even if the variables are subtracted**, i.e. the errors do not cancel each other out. (Actually, the standard deviations of independent random variables are related to each other like the lengths of orthogonal vectors from which the length of the resultant vector is calculated according to the multidimensional version of the Pythagorean theorem.)

Standardization. Standardization is a special application of the shifting and scale-changing transformations:

$$Y=\frac{X-E(X)}{D(X)}=\frac{X-\mu}{\sigma}\,. \tag{48}$$

The standardized random variable Y obtained this way has 0 for its expected value and 1 for its standard deviation as well as for its variance.

Standardization usually leads out from the distribution family of X. For instance, if X is distributed as Poisson, then Y is certainly not. However, if X happens to be $N(\mu,\sigma^2)$ normal, then Y is an $N(0,1)$ standard normal random variable. For instance, the left panel in FIGURE 1 makes use of standardization to show the properties of one-dimensional normal distributions. As standardization shifts the distribution such that the mean becomes 0, this transformation applied to the exponential distribution on the right panel in the same figure would result in a shifted exponential, which however is not considered an exponential distribution. (The exponential distribution is kind of 'attached to the soil' as its mode is at $t = 0$ by definition.)

2.2. The central limit theorem

Let X be a random variable with any distribution, having expected value μ and standard deviation σ. Let X_i denote the random variable representing the ith result of n independent trials to observe X. (The random variables X_i are obviously independent of each other and they have the same distribution as X.) Then the sample mean $\overline{X} \equiv (X_1+X_2+\ldots+X_n)/n$ is asymptotically $N(\mu, \sigma^2/n)$ normal (as a random variable).

The practical content of the central limit theorem is as follows. If we want to figure out the 'exact' value of a physical quantity by calculating the average of measured data, then (1) the estimate thus obtained is still not completely accurate. However, (2) the accuracy is better than that of the original data (it is clear: otherwise the averaging would not make any sense at all). Moreover it is also reasonable to assume that (3) the somewhat erroneous estimates that we could get for the 'exact' value in this way are normally distributed about the said 'exact' value, no matter what the original data distribution was like (e.g. it could be uniform or exponential or anything provided that the expected value and the variance exist).

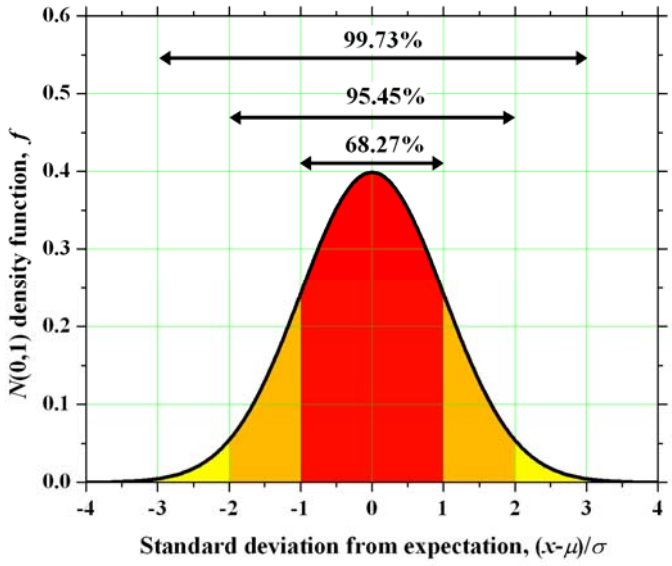

FIGURE 2. Normal distribution is a well studied distribution. If we have a reason to believe (e.g. our belief is based on the central limit theorem) that a distribution is normal (or Gaussian as often called) we can foretell that ~68% of the data is supposed to be within $\mu \pm \sigma$, where the mean μ is estimated by Eq. (29) and the standard deviation σ by Eq. (32), for instance. We can also see that ~95.5% of the data should be within $\mu \pm 2\sigma$, and less than 0.3% is expected to be outside the $\mu \pm 3\sigma$ limit.

The central limit theorem makes normal distribution one of the most important continuous distributions. Also, the fact that we know a lot about normal distribution makes the central limit theorem a very practical thing, namely, whenever it is in action, it is easy to judge how large deviations are to be expected from a given mean and how often a given deviation is likely to occur (see FIGURE 2).

Note that the central limit theorem can be phrased not only for averages but also for sums (see FIGURE 3) and, moreover, with much weaker conditions than specified above.

For n-sums, e.g., the central limit theorem can be phrased like this: Under the conditions specified above, the sum $\Sigma_n \equiv (X_1+X_2+\ldots+X_n)$ is asymptotically $N(n\mu,n\sigma^2)$ normal.

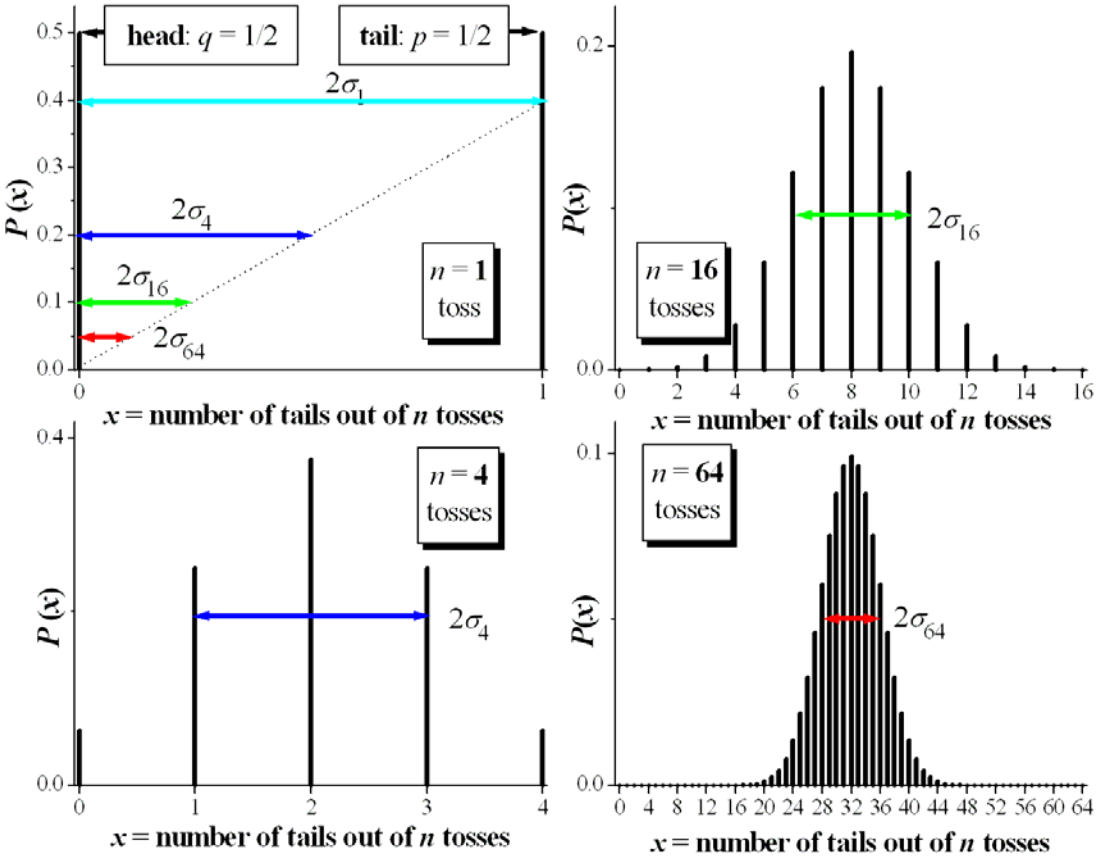

FIGURE 3. The above sequence of figures shows the distributions of the possible outcomes of four series of coin tossing—four series of Bernoulli trials—each representing one particular $B(n,p)$ binomial distribution (see later on). The probability of heads/tails has been fixed to $q = p = 0.5$, and the number of tosses (n = 1, 4, 16, 64) has been increased by a factor of 4 in the subsequent Bernoulli series. The figures demonstrate several things at the same time. (1) **Central limit theorem**: The sum of a large enough number of independent random variables having a common distribution (such as the total number of tails obtained in a long series of single tosses) is normally distributed. (2) The binomial distribution can be approximated by normal distribution if n is large enough. (3) According to the upper left figure, **if the number of data is quadrupled, then the accuracy of the average will double**. (Note that the $2\sigma_n$ arrows copied from the rest of the graphs halve as n gets 4 times larger.) The average comes into the picture because the ratio x/n (lying between 0 and 1) can also be interpreted as the average of n Bernoulli random variables that can only assume the values 0 and 1 (see later on). (4) **Law of large numbers**: As n increases, the relative width of the binomial distribution gradually diminishes, which guarantees that the relative frequency x/n observed will eventually match the probability $p = 0.5$ as $n \to \infty$. (Note that the ratio x/n represents not only the average value of the Bernoulli variable, but also the **relative frequency** of tails in a concrete Bernoulli series.)

2.3. Convolution and deconvolution

The convolution and the sum of independent random variables. If X and Y are independent (continuous) random variables with density functions f_X and f_Y, as well as distribution functions F_X and F_Y, respectively, then the density function f_{X+Y} and the distribution function F_{X+Y} of the random variable $Z = X+Y$ are given by the following convolutions:

$$f_{X+Y}(z) = f_X * f_Y(z) \equiv \int_{-\infty}^{+\infty} f_X(z-u) f_Y(u) \mathrm{d}u \,, \tag{49}$$

$$F_{X+Y}(z) = F_X * F_Y(z) \equiv \int_{-\infty}^{+\infty} F_X(z-u)\, \mathrm{d}F_Y(u) = \int_{-\infty}^{+\infty} F_X(z-u) f_Y(u)\, \mathrm{d}u \,. \tag{50}$$

The integral determining f_{X+Y} is called the convolution of density functions, that determining F_{X+Y} is referred to shortly as the convolution of the distributions. (The latter can also be calculated when X is discrete with a finite number of spectrum points. Then, in the case of continuous distributions, the density function f_{X+Y} can be obtained by differentiation if necessary.) The convolution as an operation is commutative, just like the addition. If the distributions are such that $f_i(u) = 0$, for $u \le a_i$, then:

$$f_1 * f_2(z) = \int_{a_2}^{z-a_1} f_1(z-u) f_2(u) \mathrm{d}u \,. \tag{51}$$

If therefore $a_1 = a_2 = 0$, then the integration proceeds between 0 and z.

(#8) Uniform distribution: *U*(*a*, *b*). In the following discussion we will sometimes refer to the (continuous version of the) uniform distribution that is characterized by a rectangular density function between the limits a and b. More precisely, the **density function** of the $U(a, b)$ uniform distribution is $f(x) = 1/(b-a)$ in the interval (a, b) and $f(x) = 0$ elsewhere. Its **expected value** is $\mu = (a+b)/2$ and its **variance** is $\sigma^2 = (b-a)^2/12$. The density function of the uniform distribution $U(0, 1)$ is shown in FIGURE 4.

Deconvolution. Making use of the property of the characteristic function (Fourier transform) expressed by Eq. (24), we can find a simple solution for expressing one of the components from a (density) function having the form of a convolution. We can write:

$$\varphi(f_{X+Y}) = \varphi(f_X * f_Y) = \varphi(f_X)\ \varphi(f_Y) \,. \tag{52}$$

Therefore the density function of X, e.g., can be expressed by the **inverse Fourier transform**:

$$f_X = \varphi^{-1}\left(\frac{\varphi(f_X * f_Y)}{\varphi(f_Y)} \right). \tag{53}$$

(#9) The significance of convolution/deconvolution exceeds the boundaries of probability theory. Some nuclear spectra can also be described by a convolution-type integral. For instance, a transmission *Mössbauer spectrum* is the convolution of two functions one of which is characteristic of the source of radiation while the other of the absorber (sample). The latter function contains all the parameters that the spectroscopist can be interested in, which explains why some methods of spectrum evaluation include the calculation of deconvolution as well.

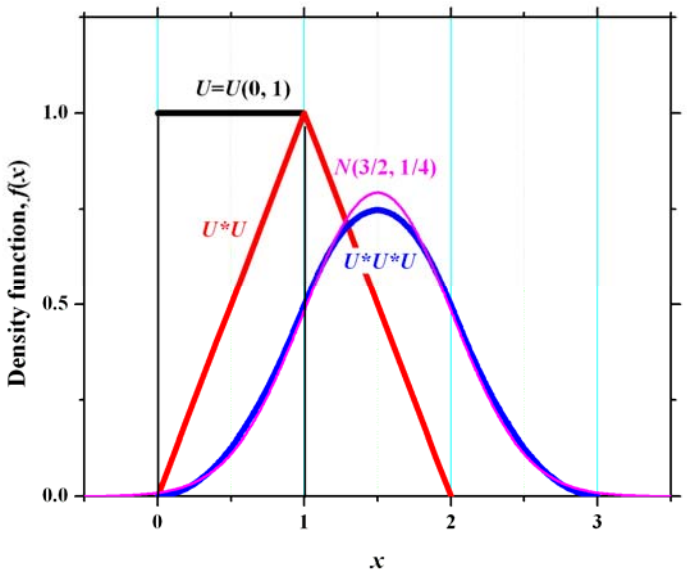

FIGURE 4. This figure shows the rectangular density function of the uniform distribution $U(0, 1)$ together with its convolutions. Generating $U(0, 1)$ (pseudo) random numbers is a standard feature of many 'scientific' calculators. The convolution power $U*U$ represents the distribution of the sum of two such numbers. Similarly, if three such numbers are added, the distribution of the sum will be $U*U*U$. The example demonstrates how soon central limit theorem can be caught in the act even in the case of a distribution far from being bell shaped. Note that the density function of $U*U*U$ is composed of parabolic fragments but still it is rather close in shape to that of a normal distribution of the same mean and variance.

2.4. Random sums

Sums of a random number of random variables—in short: **random sums**—are often encountered at several stages of radiation detection. The problem usually presents itself disguised as a product:

Consider about μ_N items, each consisting of about μ_X parts (or, alternatively, each weighing/measuring/worth etc. about μ_X kg/m/dollars etc.). Obviously the whole lot of parts (mass/length/value etc.) adds up to about $\mu_N\mu_X$ (kg/m/dollars etc.). The question to be answered is the following. **What is the error of this 'product'?**

The answer is given by the following theorem (Feller 1968, Korn and Korn 1968). This theorem gives the solution for the recurring problem of **Bernoulli sampling** that will be discussed later on.

Let $X_1, X_2, \ldots X_N$ be a sequence of mutually independent (integral valued) random variables with a common distribution. Let N be itself an integral valued variable, being independent of the others. Let $\Sigma_N = X_1+X_2+\ldots+X_N$. Let $G_X(s)$ denote the (common) generating function of the variables $X_1, X_2, \ldots X_N$. Let $G_N(s)$ be the generating function of N, and $G_\Sigma(s)$ that of the random sum Σ_N. Then we have:

$$G_\Sigma(s) = G_N(G_X(s))\,. \tag{54}$$

Applying the chain rule of differentiation to Eq. (54), we can calculate the expected value and the variance of the random sum from Eqs. (21):

$$\mu_\Sigma = \mu_N\mu_X\,; \qquad \sigma_\Sigma^2 = \mu_X^2\sigma_N^2 + \mu_N\sigma_X^2 \tag{55}$$

Thus the square of the relative deviation of the random sum is:

$$\left(\frac{\sigma_\Sigma}{\mu_\Sigma}\right)^2 = \left(\frac{\sigma_N}{\mu_N}\right)^2 + \frac{1}{\mu_N}\left(\frac{\sigma_X}{\mu_X}\right)^2. \tag{56}$$

This asymmetrical result implies that if the (expected) number of terms is large enough, then the variance of the random sum is principally determined by the variance of the number of these terms (i.e. the variance of the individual terms is of relatively less importance). It is worthwhile to compare this result with Eq. (61), showing the formula of error propagation for 'real' products.

2.5. General approximation of error propagation

For any sequence of random variables, the following approximation holds for the error of the formula $f(X_1, X_2,\ldots X_n)$ calculated from these variables:

$$D^2(f) \approx \sum_{i=1}^{n}\left[\left(\frac{\partial f}{\partial x_i}\right)^2_{\boldsymbol{x}=\boldsymbol{\mu}} D^2(x_i)\right] + 2\sum_{i<j}^{n}\left[\left(\frac{\partial f}{\partial x_i}\right)_{\boldsymbol{x}=\boldsymbol{\mu}}\left(\frac{\partial f}{\partial x_j}\right)_{\boldsymbol{x}=\boldsymbol{\mu}} \mathrm{Cov}(x_i, x_j)\right]. \tag{57}$$

If the variables are mutually **independent**, then the second sum containing the covariances becomes zero and therefore we have:

$$D^2(f) \approx \sum_{i=1}^{n}\left[\left(\frac{\partial f}{\partial x_i}\right)^2_{\boldsymbol{x}=\boldsymbol{\mu}} D^2(x_i)\right]. \tag{58}$$

The subscript $\boldsymbol{x} = \boldsymbol{\mu}$ is to indicate that the derivatives must be calculated at the expected values of the variables. (In practical terms, the expected value is the 'exact' value around which the 'measured' value of the variable fluctuates.)

Equations (57)-(58) have been obtained from the (first-order) differential—i.e. from a linear approximation—of the function $f(X_1, X_2,\ldots X_n)$. Therefore the related formulae are quite general in the sense that f can be any type of differentiable function, however the approximation can only be used for continuous variables and for relatively small errors. On the other hand, Eqs. (46)-(47) obtained for linear expressions—actually the special cases of Eqs. (57)-(58)—are exact.

2.6. Formulae for products/ratios of independent random variables

For the **expected value of products/ratios** we have the following exact formula:

$$E\left(\frac{XY}{Z}\right) = E(X)\,E(Y)\,E\left(\frac{1}{Z}\right). \tag{59}$$

If $D^2(Z)/E^3(Z)$ is small enough, then we get the following approximation by expanding the expression $1/Z$ into Taylor series about $E(Z)$:

$$E\left(\frac{XY}{Z}\right) \approx \frac{E(X)\,E(Y)}{E(Z)}. \tag{60}$$

The **approximation of standard deviation of products/ratios** can be obtained from the appropriate form of Eq. (58):

$$\left[\frac{D\left(\frac{XY}{Z}\right)}{E\left(\frac{XY}{Z}\right)}\right]^2 \approx \left[\frac{D(X)}{E(X)}\right]^2 + \left[\frac{D(Y)}{E(Y)}\right]^2 + \left[\frac{D(Z)}{E(Z)}\right]^2 . \tag{61}$$

We can see that, in the case of multiplication and division, it is the squares of the relative deviations that add up, rather than the variances. Note that the last term representing division is also added rather than subtracted in the above formula.

We have seen earlier that Eq. (56) gives a different error formula for a different type of 'product'. In the case of products therefore it is worthwhile to think it over whether or not a random sum is hiding behind the problem. This can only happen, of course, when one of the variables is an integer having no dimension.

3. Special Distributions

On the next few pages we will go through the distributions playing a role in nuclear measurements as well as in the evaluation of nuclear spectra.

3.1. Bernoulli distribution: $B(1, p)$

Mass function: ..... $P(X = x; p) = \begin{cases} p & \text{if } x = 1 \\ q & \text{if } x = 0 \end{cases}$ where: $q = 1 - p$

Generating function: $G(s) = q + ps$

Expected value: p ..... **Mode**: $\begin{cases} 1 & \text{if } p \geq q \\ 0 & \text{if } p \leq q \end{cases}$

Variance: .....pq **Relative deviation**: $\sqrt{\frac{q}{p}}$

Some authors consider the terms Bernoulli distribution and binomial distribution synonyms. We will reserve the former for the simplest variant of binomial distribution.

Interpretation. Consider a dichotomous experiment that has only two possible outcomes (alternative events) like a 'coin toss'. One of the alternative outcomes (e.g. 'heads') is generally called 'success' while the other (e.g. 'tails') 'failure'. The probability of 'success' is denoted by p, that of 'failure' by $q = 1 - p$. Now, the Bernoulli variable X is defined as follows:

$$X(\text{'success'}) \equiv 1; \quad X(\text{'failure'}) \equiv 0\,. \tag{62}$$

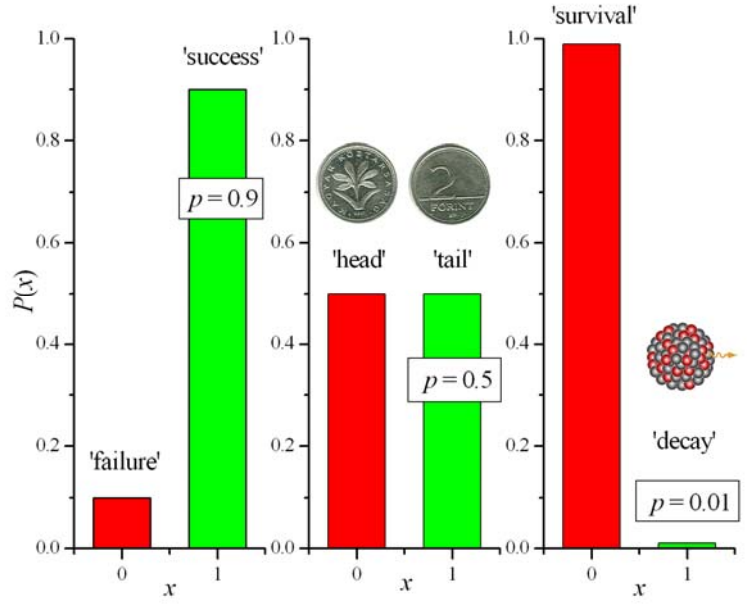

FIGURE 5. Three basic types of **Bernoulli distribution** classified according to the ratio of p ('success' ⇔ 1) to q ('failure' ⇔ 0). Any series of Bernoulli trials results in **binomial distribution** (see FIGURE 3), however only the third type, characterized by a low probability of success, leads to **Poisson distribution**. **Radioactive decay** usually belongs to the latter category.

(#10) The repeated independent trials associated with such a dichotomous ‘game’ are called **Bernoulli trials**.

(#11) As regards nuclear methods, another dichotomous game comes to mind: the fate of a **radioactive nucleus** in a period of time. The alternative outcomes are: ‘decay’ (characterized by probability p) and ‘survival’ (characterized by probability q).

(#12) $B(1, p)$ Bernoulli random numbers can be easily obtained from $U(0, 1)$ random numbers uniformly distributed in the interval (0, 1). (As mentioned before, such random numbers are readily available using ‘scientific’ calculators.) The algorithm is straightforward: we choose $X = 1$, whenever the $U(0, 1)$ uniform random number happens to be less than p, and we choose $X = 0$ otherwise.

3.2. Binomial distribution: $B(n, p)$

Mass function: ..... $P(X = x; n, p) = \binom{n}{x} p^x q^{n-x}$ $(x = 0, 1, \ldots n)$ $q = 1 - p$

Generating function: $G(s) = (q + ps)^n$

Expected value: $\mu \equiv np$ .....**Mode**: $\begin{cases} \mu - q \\ \mu + p \end{cases}$ and if $\mu + p$ is integer; $[\mu + p]$ if $\mu + p$ is not integer

Variance: ..... $npq = \mu q$ **Relative deviation**:... $\sqrt{\frac{q}{\mu}} = \sqrt{\frac{q}{pn}}$

Interpretation. Consider a dichotomous 'game'. Let p denote the probability of 'success' in a single trial. Let us perform a series of n Bernoulli trials. Let X_1, X_2, … X_n denote the independent Bernoulli variables belonging to the respective trials. Then the random variable $X \equiv X_1 + \ldots + X_n$ has a $B(n, p)$ binomial distribution.

(#13) Note that X means the number of successful outcomes in the series of n Bernoulli trials. Note also that the above interpretation justifies the use of the symbol $B(1, p)$ for the Bernoulli distribution.

(#14) The above remark provides a straightforward recipe for the simulation of $B(N, p)$ random numbers if we have a $U(0, 1)$ random number generator. All we need to do is generate N random numbers of $B(1, p)$ Bernoulli distribution according to remark **(#12)**. The sum of the N Bernoulli numbers will then result in a $B(N, p)$ random number.

(#15) The coin-toss game can be performed either by tossing one single coin n times, or by tossing n coins simultaneously. If the only question is the probability of tossing exactly x 'heads' out of n tosses, then the two games are essentially identical, and the answer is provided by the mass function of the $B(n, p)$ binomial distribution. (It follows from symmetry that by swapping the roles we get a similar formula for the number of 'tails', too.)

(#16) Referring to remark **(#11)**, the simultaneous 'coin-toss game' can be paraphrased like this: Let us consider n identical radioactive nuclei, any of which decays with the same probability p during a certain period of observation. What is the probability that exactly x nuclei will decay? The answer is given by the mass function of the $B(n, p)$ distribution. In other words, the fluctuation of the number of nuclei that actually decay over a given period of time can be described by the binomial distribution. (It follows from symmetry that by swapping roles—let p denote the probability of 'survival', and x the number of nuclei escaping decay over the same period—we get the same distribution for the number of 'survivors'.)

(#17) Let n denote the number of radionuclei like before. Let p represent the probability that a selected nucleus will decay within a given period of time, and η the probability that the decay of a nucleus will be actually observed with the given detector system. Then $p\eta$ obviously measures the probability that a given nucleus decays over the period of time *and* this decay will be actually observed. What is the probability that we will observe the decay of exactly x nuclei? The answer is given by the mass function of the $B(n, p\eta)$ distribution. In other words, the fluctuation of the number of counts measured (without background) over a given period of time can be described by the binomial distribution.

We will refer to this result as the **Bernoulli sampling of binomial distribution**, because the problem can also be presented in the following way: The number of decayed nuclei have a $B(n, p)$ distribution. Let X denote the actual number of nuclei that have decayed over the given

period of time. Let us perform a $B(1,\eta)$ sampling on the decayed nuclei, i.e. we go over them one after the other and either accept them with probability η (meaning that the decay of the nucleus has been detected) or discard them with probability $1-\eta$ (the decay of the nucleus has escaped detection). The question is how many decay events have been detected altogether and how these detected numbers are distributed in a whole series of like experiments. Thinking it over, we have a series of Bernoulli trials here, the length of which (X) itself is a random variable with a $B(n,p)$ distribution. Therefore what we should do is add up X random variables having $B(1,\eta)$ Bernoulli distribution. Using Eq. (54) obtained for random sums we find that this sum—i.e. the number of decay events actually observed or, which is the same, the number of counts—has $B(n,p\eta)$ distribution, where p is the probability of decay and $\eta < 1$ is the **probability/efficiency of detection**.

(#18) We can check for ourselves that for large values of n, i.e. for large μ, the mode(s) of binomial distribution is (are) practically equal to the expected value, which means that the expected value doubles as the most probable value. On the other hand, the relative deviation rapidly decreases with increasing n or μ. This means that the distribution—obeying the law of large numbers—is gradually 'shrinking' on the expected value (see FIGURE 3).

Owing to this, if we have measured a large enough number of counts, that single number can be considered as a fairly good estimate of the expected value. (This statement is not at all true for any distribution.) Of course, in the case of such a one-point estimate it is impossible to find out whether or not the measured data point squares with the assumption that the equipment has been faultless.

3.2.1. Properties

Addition theorem. If X_1 and X_2 are independent random variables with $B(n_1,p)$ and $B(n_2,p)$ distribution, respectively, then the random variable $X = X_1+X_2$ has $B(n_1+n_2,p)$ distribution.

(#19) As regards the example mentioned in remark **(#16)**, the above property means that if we have two 'heaps' of identical radionuclei, then not only the separate counts have binomial distribution (with expected values n_1p and n_2p as well as variances n_1pq and n_2pq, respectively), but the total number of counts as well [with expected value $(n_1+n_2)p$ and variance $(n_1+n_2)pq$]. Choosing one second as the time of observation, the addition theorem expresses the **additive property of activity**.

Normal approximation (DeMoivre-Laplace limit theorem). It follows from the interpretation as well as from the central limit theorem that for large enough values of npq ($npq \geq 6$ suffices already) the binomial distribution can be approximated by a normal distribution with expected value $\mu = np$ and variance $\sigma^2 = npq$ (see FIGURE 3 and FIGURE 6):

$$B(n,p) \approx N(np, npq)\,. \tag{63}$$

Poisson approximation. For small enough values of p and large enough values of n ($p \leq 0.1$ and $n \geq 20$ will do) the binomial distribution can be approximated by a Poisson distribution with parameter (expected value) $\mu = np$ (see FIGURE 5 and FIGURE 7):

$$B(n,p) \approx \Pi(np)\,. \tag{64}$$

(#20) Let us return to the example in remarks **(#16)-(#17)**. If the observation period is very short compared to the half-life of the radionuclide in question (i.e. $p << 1$) and if the total number of nuclei observed is large enough, then the number of decay events/counts has a

Poisson distribution. This serves with a simple 'recipe' for the estimation of the standard deviation:

$$s \approx \sigma = \sqrt{\mu} \approx \sqrt{\overline{X}} \ . \tag{65}$$

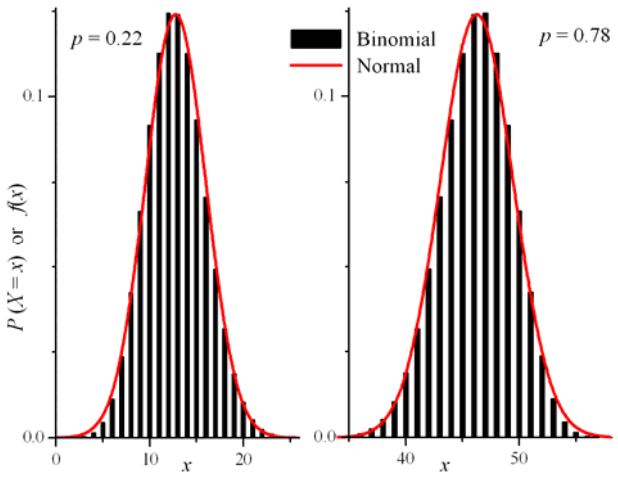

FIGURE 6. The $N(np, npq)$ normal distribution provides a rather good approximation for the $B(n, p)$ binomial distribution irrespective of the value of p provided that $npq \geq 6$. In the example shown n is merely 59, and npq is only 10, however the 'fit' is quite good already.

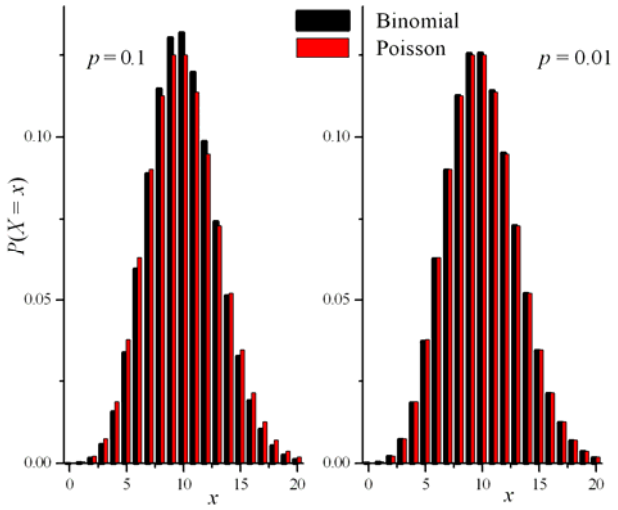

FIGURE 7. The $B(n, p)$ binomial distribution can be approximated with the $\Pi(np)$ Poisson distribution even for small values of $\mu = np$ provided that the probability of success in a single trial is small ($p \leq 0.1$). We can see that the 'fit' is reasonably good at $p = 0.1$, although the value of μ is only $np = 10$. For $p = 0.01$, on the other hand, the 'fit' looks almost perfect (see FIGURE 5).

(#21) Continuing the previous remark: If the number n of the radionuclei is large enough, which is usually the case, then the number of decay events/counts is normally distributed. Therefore the expected fluctuation of counts, e.g., can be judged by the 'rules' of normal distribution (see Table 3). For instance, the counts X_i collected over the same length of time are supposed to lie in the error intervals '$\mu \pm 1\sigma$', '$\mu \pm 2\sigma$', and '$\mu \pm 3\sigma$' with the following

frequencies (provided that the total time needed for the series of measurements is short enough compared to the half-life of the radionuclide):

$$\overline{X} \pm \sqrt{\overline{X}}: \quad 68\%; \quad \overline{X} \pm 2\sqrt{\overline{X}}: \quad 95.4\%; \quad \overline{X} \pm 3\sqrt{\overline{X}}: \quad 99.7\%. \tag{66}$$

3.3. Poisson distribution: $\Pi(\mu)$

Mass function: ..... $P(X = x;\mu) = \dfrac{\mu^x}{x!}\mathrm{e}^{-\mu} \quad (x = 0, 1, 2, \ldots)$

Generating function: ..... $G(s) = \mathrm{e}^{\mu(s-1)}$

Expected value: μ .....**Mode**: $\begin{cases} \mu - 1 \\ \text{and} \\ \mu \end{cases}$ if μ is an integer; $[\mu]$ if μ is not an integer

Variance: ..... μ **Relative deviation**: ..... $\dfrac{1}{\sqrt{\mu}}$

Interpretation. Consider a recurring instantaneous rare event (like the arrival of a background pulse) which repeats itself at random over time. We wish to count such events over a certain observation period ΔT. We will assume that the following general conditions are satisfied:

1, The outcome of the observation does not depend on the location of ΔT along the time axis.

2, The occurrences of the event in separated time intervals are stochastically independent.

We will also assume that the following '**rarity**' and '**time-proportionality**' conditions are fulfilled in the case of very short periods of time $\Delta t << \Delta T$:

3, If the observation period ΔT is divided to n equal intervals—'time cells'—of length $\Delta t = \Delta T/n$, then the majority of such intervals would not contain any event at all, while the rest contain exactly one event each, provided that n is large enough ($n \to \infty$). That is, the possibility of events appearing in doublets etc. is excluded if the 'time cells' considered are small enough.

4, At the above fine time-scale the (single) occurrence of the event within a 'time cell' is proportional to the length Δt of the interval:

$$p = P(1) = \frac{\mu}{\Delta T}\Delta t = \frac{\mu}{n}. \tag{67}$$

If the above assumptions are satisfied, then the total number of events observed (i.e. the number of counts X) has a $\Pi(\mu)$ Poisson distribution.

(#22) Condition **4** can be better understood, if μ is regarded as the expected number of events occurring over the observation period ΔT. (According to the data in the box of characteristic parameters, this interpretation is justified, because $E(X) = \mu$ for the Poisson distribution.) Thus the ratio μ/n equals the average number of events occurring in a single 'time cell'. Note that as n approaches infinity, this average sooner or later drops below 1, however large the value of μ is (e.g., $\mu/n \approx k/m$, where $k < m$). Owing to condition **3** this means that—asymptotically—on an average k cells out of m contain exactly one event each, while the rest are empty. In other words, the probability that a cell is 'filled' with an event is k/m ($\approx \mu/n$), i.e. condition **4** is satisfied.

(#23) The above conditions are fulfilled, e.g., in the case of the background caused by cosmic radiation etc. Therefore **background counts** observed over a given period of time have Poisson distribution.

(#24) We can see that, for large values of μ, the mode(s) is(are) practically equal to the expected value, i.e. the expected value doubles as the most probable value in the case of the

Poisson distribution. On the other hand, the relative deviation rapidly decreases as μ increases. Thus the distribution is gradually 'shrinking' on the expected value.

Consequently, if we measure a large enough background, then this single measurement can provide a fair estimate for the expected value, the square root of which can be used for the estimation of the standard deviation. (See also remark **(#18)** on the unreliability of such one-point estimates.)

3.3.1. Properties

Bernoulli sampling. Consider a random variable with a $\Pi(\mu)$ Poisson distribution, which counts the occurrences of certain types of events (e.g. events of radioactive decay). Let us perform a $B(1,\eta)$ Bernoulli sampling on the counted events, meaning that we either accept them with probability η, or discard them with probability $(1-\eta)$. For instance, η can indicate the efficiency of detection, in which case 'acceptance' means that the decay event in question has been actually detected. It follows from Eq. (54) that the number of events passing such a sampling (e.g. the number of detector pulses caused by the radiation of decaying nuclei) has a $\Pi(\mu\eta)$ Poisson distribution. (See also remark **(#17)**.)

Addition theorem. If X_1 and X_2 are independent random variables with Poisson distributions $\Pi(\mu_1)$ and $\Pi(\mu_2)$, respectively, then the random variable $X = X_1+X_2$ has a $\Pi(\mu_1+\mu_2)$ Poisson distribution.

(#25) As a consequence of remark **(#20)** as well as due to the addition theorem the following conclusion can be drawn: Suppose that we observe the decay events in a specimen containing a **mixture** of two (or more) radionuclides. If the observation period is short enough in comparison with any of the half-lives (i.e. any single nucleus will survive the period with a high probability), then the total number of decay events has a Poisson distribution. It follows from the property of Bernoulli sampling that the same applies to the total number of counts.

(#26) Continuing the previous remark as well as remark **(#23)**: The number of counts measured with background also has a Poisson distribution.

Normal approximation. It follows from the addition theorem and from the central limit theorem that for large enough μ the Poisson distribution can be approximated by a normal distribution (see FIGURE 8):

$$\Pi(\mu) \approx N(\mu, \mu)\,. \tag{68}$$

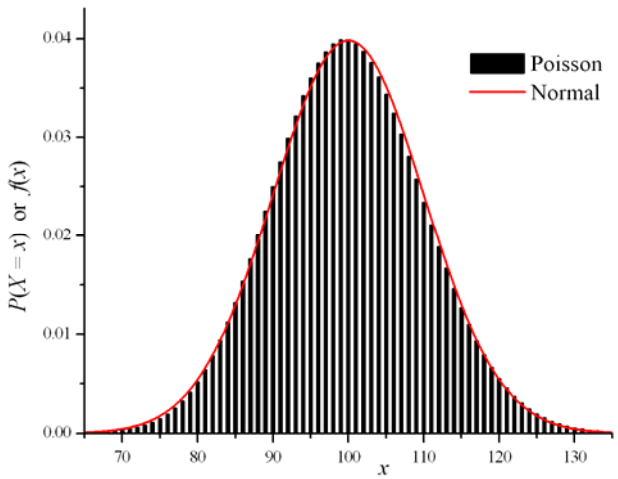

FIGURE 8. The $N(\mu, \mu)$ normal distribution provides an almost perfect fit for the $\Pi(\mu)$ Poisson distribution at as low a parameter value as $\mu = 100$.

(#27) Continuing the previous remarks: If the counts measured together with background are large enough, then their fluctuation can be judged by the ‘rules’ of the normal distribution in the same way as suggested in remark **(#21)**. See Eq. (66) and Table 3.

3.4. Exponential distribution: $\gamma(1, \lambda)$

Density function: ..... $f(t) = \begin{cases} \lambda \mathrm{e}^{-\lambda t} & \text{if} \quad t > 0 \\ 0 & \text{if} \quad t \le 0 \end{cases}$

Distribution function: ..... $F(t) = 1 - \mathrm{e}^{-\lambda t}$

Characteristic function: ..... $\varphi(u) = (1 - \mathrm{i}\, u / \lambda)^{-1}$

Expected value: .. $\tau \equiv \dfrac{1}{\lambda}$ **Median**:..... $T_{1/2} \equiv \dfrac{\ln 2}{\lambda} \approx \dfrac{0.7}{\lambda}$

Variance: ..... $\dfrac{1}{\lambda^2}$ **Relative deviation**:..... 1

Interpretation. Consider an **'ageless' but 'mortal' entity** the lifetime of which is characterized by the random variable $T \in (0,\infty)$. Agelessness means that—at any moment of its life—the life expectancy of such an entity is independent of its age, i.e. of the time it has already lived up to that point. Therefore if we find it alive at the moment $t > 0$ (i.e. if $T > t$), then the (conditional) probability of still finding it alive at a later moment $t+s$ only depends on the time elapsed since the moment t (i.e. it only depends on s):

$$P(T > t + s \,|\, T > t) = P(T > s)\,. \tag{69}$$

Let us further assume that the probability that the entity's 'death' (or 'decay' as is usually called) occurs within a given time interval is proportional to the length Δt of that interval, provided that Δt is small enough. In other words, $P_{\text{decay}} = 1 - P(T > \Delta t) = \lambda\, \Delta t$, where λ is the **decay constant**.

If the above conditions are satisfied, then the lifetime T is a random variable with $\gamma(1, \lambda)$ exponential distribution.

(#28) It is obvious that the above conditions have been 'tailored' to fit the disintegration of radioactive nuclei. Therefore we can declare that **the lifetime distribution of radionuclei is exponential**. The same is true for the excited states of nuclei as well as atoms.

There are cases when the exponential distribution is related to waiting times between random events rather than to lifetimes. In such cases the adjective '**ageless**' is swapped for '**memoryless**' (Goodman 1988). Anyway, both metaphors are verbal expressions for the same mathematical condition and as such they determine the same (exponential) distribution.

Another phenomenon related to exponential distribution is the **absorption (attenuation) of gamma photons** in a homogeneous absorber. The correspondence between the roles is as follows: $t \Leftrightarrow d$ (depth of penetration), $\lambda \Leftrightarrow \mu_1$ (linear attenuation coefficient), $T_{1/2} \Leftrightarrow D_{1/2}$ (halving thickness), $\tau \Leftrightarrow \delta$ (mean free path). Agelessness is translated now to **indefatigability** meaning that the photons—in contrast to ionizing particles—are not losing energy in a continuous way as they proceed in the absorber. However, being mortal, they may 'drop dead' suddenly via photoelectric effect, for instance.

(#29) The distribution function (see the red curve in FIGURE 9) expresses the monotonically increasing probability that the nucleus will decay sometime over the period $(0, t)$. The 'mortality' of the atom is expressed by the normalization condition $F(\infty) = 1$ (which is fulfilled). The fact, on the other hand, that the atom was actually found 'living' at the moment 0 corresponds to the condition $F(0) = 0$ (also fulfilled). Thus the probability that the atom will survive t is:

$$1 - F(t) = e^{-\lambda t} . \tag{70}$$

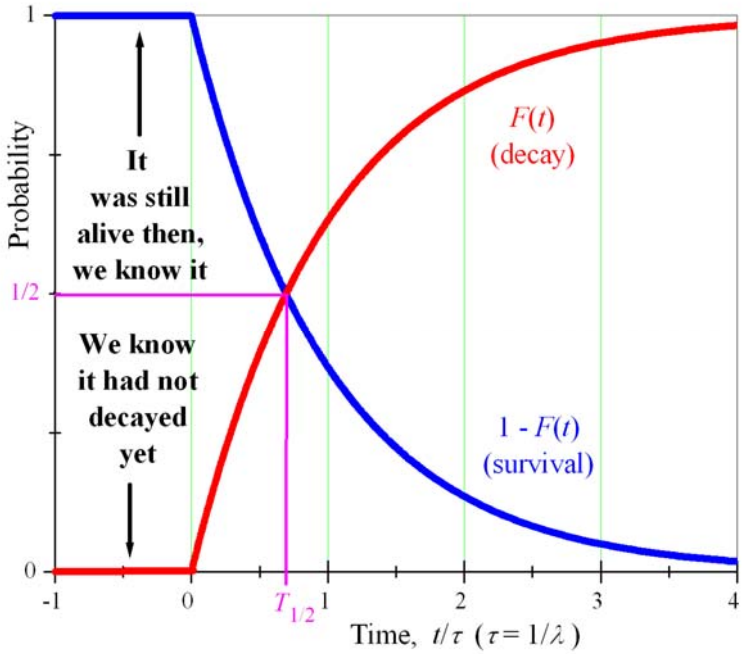

FIGURE 9. The **exponential distribution function** $F(t)$ (red curve) expresses the probability that the 'entity'—looking at it from the present moment 0—will be 'dead' by the moment t (i.e. that it will decay somewhere between 0 and t). The dotted curve, on the other hand, shows the probability of the complementary event, i.e., that the same 'entity' will survive the period $(0,t)$. For $t > 0$ the curve of the **exponential density function** is obtained from the blue curve by multiplying the latter with λ. (However, for $t \le 0$, the density function is zero, not 1, as we can see in FIGURE 1.)

(#30) Using the above equation, one can get **exponentially distributed random numbers** from uniformly distributed ones having $U(0, 1)$ distribution. If, namely, the random number/variable X is uniformly distributed in the interval $(0, 1)$, then the random number/variable T calculated from the formula

$$T = -\frac{\ln X}{\lambda} \tag{71}$$

will have a $\gamma(1,\lambda)$ exponential distribution (Goodman 1988, Lux and Koblinger 1991).

(#31) In the case of the lifetime distribution of radionuclides, excited states etc., the expected value τ is called the **mean life**[7], while the median $T_{1/2}$ is referred to as the **half-life**. The explanation for the name 'half-life' is given in the next section on the exponential law of radioactive decay. We will see that the half-life is independent of the time elapsed, which is an obvious proof of the agelessness of radionuclides.

(#32) Looking at the location parameters of exponential distribution, we can see that the rule of thumb expressed by Eq. (13) works fairly well for the mode (0), median ($T_{1/2}$) and mean (τ), the ratio of their respective distances being 0.7:0.3 (= 2.1:0.9). This ratio is almost equal to the predicted ratio 2:1 (see also the upper right diagram in FIGURE 1).

[7] Note that physicists often use the term **lifetime** not only in the sense we do here, but also in the sense **mean life**. Fortunately, in the really important cases, i.e., when quantitative statements are made (e.g., 'the lifetime of the radionuclide is 10 s'), the ambiguity is removed and the reader can be sure that such a statement actually refers to the mean life.

(#33) The relative-deviation formula ($\sigma_{\text{rel}} = 1$) explains why it is practically hopeless to give a usable **mean-life estimate** from the observation of the decay of a single atom. (Remember, however, that in the case of the Poisson distribution a large enough observed value could be considered as a fairly good estimate of the expected value.) On the other hand, if we change the time-scale to logarithmic, e.g., by one of the following transformations:

$$w = \ln t \quad \text{or} \quad u = \ln \frac{t}{\tau} \tag{72}$$

(but decimal logarithm will also do), then the density function of the new variable (readily obtained from the exponential distribution function by differentiation using the chain rule):

$$f(w) = \frac{1}{\tau} \exp\left(w - \frac{\mathrm{e}^w}{\tau} \right) \quad \text{or} \quad f(u) = \exp(u - \mathrm{e}^u) \tag{73}$$

reaches its maximum at the value corresponding to the condition $t = \tau$ (Schmidt et al. 1984).

The halfwidth of the asymmetric bell-shaped curve in FIGURE 10 is rather large (FWHM ≈ 2.5 representing about a decimal order of magnitude in uncertainty), however, it can serve as a basis for estimating the mean life τ from the mode of a histogram that has been drawn using the logarithm of a few dozens of lifetime data. This method is used for the identification of super-heavy elements (SHEs) when only the decay of a couple of man-produced transactinide atoms can be observed at a time.

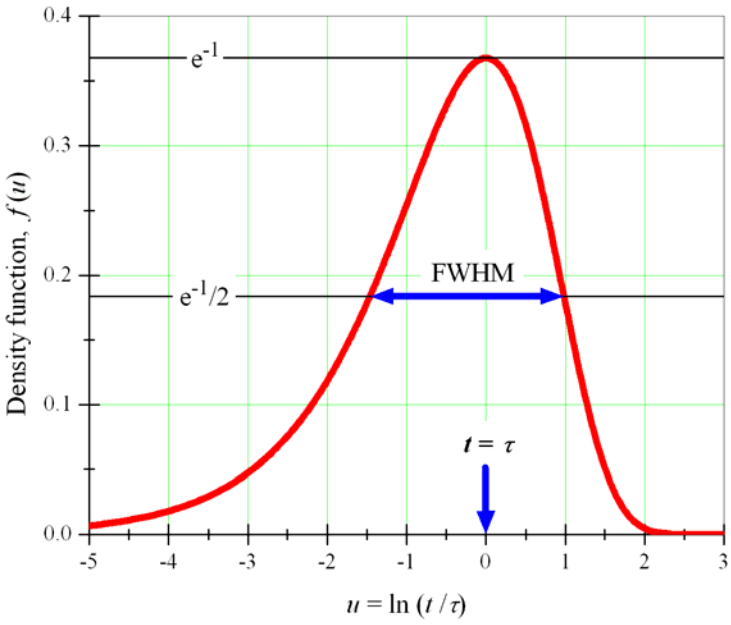

FIGURE 10. The density function of the logarithm of exponentially distributed lifetimes. Using this trick, the **estimation of the mean life** can be reduced to the determination of the mode (i.e. the most frequent value) of a logarithmized lifetime histogram related to the above distribution. This method can be used whenever the number of lifetimes that can be measured is so small that the standard method of mean life determination ($\tau = 1/\lambda$) based on the exponential fit of lifetime data fails. See also remark (#35).

3.4.1. Connection with the exponential law of radioactive decay

Suppose that at moment zero we have N_0 atoms. The question is **how many atoms will live at moment *t***. Let $N(t)$ denote this number. Now if we repeated the experiment several times starting with the same number of atoms (N_0), then the number of the survivors would vary showing that $N(t)$ is a random variable. In other words, the $N(t)$ values would fluctuate about their time-dependent mean $\langle N\rangle(t)$. For any fixed value of t, $N(t)$ has a $B[N_0, p(t)]$ binomial distribution for which the value of $p(t)$—i.e. the probability that a given atom is still intact at moment t—is calculated from Eq. (70) as explained in remark **(#29)** and FIGURE 9:

$$P\big(N(t)=k\big)=\binom{N_0}{k}\left(\mathrm{e}^{-\lambda t}\right)^k\left(1-\mathrm{e}^{-\lambda t}\right)^{N_0-k}\quad\big(k=0,1,\ldots N_0\big). \tag{74}$$

Now if the question is put this way: **How many atoms *are expected* to live at moment *t*?**, then the answer is provided by the expected value of the binomial distribution:

$$\langle N\rangle(t)\equiv E_B\big(N(t)\big)=N_0\,\mathrm{e}^{-\lambda t}. \tag{75}$$

We can easily recognize the exponential law of radioactive decay in the above formula. The correspondence $\langle N\rangle(t)\Leftrightarrow N$ reminds us that **the exponential law applies to the expected number of atoms** rather than to the concrete numbers that we measure. The latter show a fluctuation about those expectations according to the standard deviation of the binomial distribution:

$$\sigma_N(t)=D_B\big(N\big)(t)=\sqrt{N_0\,\mathrm{e}^{-\lambda t}(1-\mathrm{e}^{-\lambda t})}=\sqrt{\langle N\rangle(t)\,(1-\mathrm{e}^{-\lambda t})}\,. \tag{76}$$

Note that the above formula gives $\sigma=0$ both for $t=0$ and for $t=\infty$. This is due to the following certainties: (1) all of the N_0 atoms were intact at time 0; (2) all of them will decay sooner or later.

As for medium lengths of time (i.e. for periods that are commensurable with the mean life of the atoms), the acceptability of the approximation $\langle N\rangle(t)\approx N$ can be judged by the relative deviation formula:

$$\sigma_{\mathrm{rel}}(t)=\frac{\sigma_N(t)}{\langle N\rangle(t)}=\sqrt{\frac{1-\mathrm{e}^{-\lambda t}}{\langle N\rangle(t)}}<\frac{1}{\sqrt{\langle N\rangle(t)}}\,. \tag{77}$$

The above result conveys a reassuring message, namely, if the number of intact atoms is still large enough, then the uncertainty of their number (as expressed by the relative deviation) is negligible. That means that the deterministic expression

$$N=N_0\,\mathrm{e}^{-\lambda t}=N_0\,\mathrm{e}^{-t/\tau}=N_0\,2^{-t/T_{1/2}} \tag{78}$$

usually referred to as **the exponential law of radioactive decay** stands the test of statistics as well.

(#34) The reason for calling the median 'half-life' can be best understood from that variant of the exponential law in which the usual base e is changed to 2. The last expression in Eq. (78) shows that N_0—representing the number of atoms present at time $t=0$—will drop to $N_0/2$ by the time $t=T_{1/2}$. Then, after another half-life, the number of survivors will be halved again, and so on. Experience shows that the half-life of radionuclei is independent of both the

choice of the origin and the number of halving cycles that have already elapsed. This can be regarded as a proof for the agelessness of radionuclei. (If they were subject to aging, then the half-life should gradually decrease—an indication that the rest of the population is getting closer to the end of their lives.)

(#35) The exponential law can be written formally as:

$$\mu(t) = \mu(t; \boldsymbol{b}) = \mu(t; b_1, b_2) = b_1 \, \mathrm{e}^{-b_2 t} \, . \tag{79}$$

The above expression is an example of a model function, which fits the nuclear spectrum consisting of the spectrum points[8]

$$(t_i, N(t_i)) \quad i = 1, 2, \ldots k \tag{80}$$

at an appropriately chosen value ($\boldsymbol{a}$) of the parameter vector $\boldsymbol{b}$:

$$\boldsymbol{b} = (b_1, b_2) = (N_0, \lambda) \equiv \boldsymbol{a} \tag{81}$$

By fitting we mean that Eq. (79) will produce the expected value of $N(t_i)$ for any value of the independent variable t_i. (See Chapter 5 on fitting nuclear spectra.)

3.4.2. Exponential law in a binomial way

FIGURE 11 and FIGURE 12 help to understand the connection between exponential law on the one part as well as binomial distribution and exponential distribution on the other.

FIGURE 11 shows the possible 'fate' of 1000 atoms through 28 subsequent periods of time—time units—by taking a census of the population at the end of each period. The population numbers $N(t)$ have been simulated by a sequence of random numbers with $B(N_t, p)$ distributions[9]. (The value of the parameter p—representing the probability that a single atom will survive the next period—has been fixed to $p = 0.9$.) Thus the starting value of $N(t) \equiv N_t$ has been $N_0 = 1000$, then the number of survivors has been decreased stepwise by using the recursion algorithm $B(N_{t-1}, p) \Rightarrow N_t$.

The time-dependence of the expected number of survivors is expressed by the monotonically decreasing sequence of the binomial expected values N_0p, $(N_0p)p$, $[(N_0p)p]p$, … $N_0p^{\{t\}}$. Note that this sequence consists of the substitution values of a decreasing exponential function taken at the end of each period. The explanation for this is as follows. Since no atoms are born in any of the periods, each 'census' simplifies to a survival test. The mathematical equivalent of this is the $B(1, p)$ **Bernoulli sampling** of the $B(N_{t-1}, p)$ distribution of the previous population. Note that the exponential character of the expected value as a function of time is explained by the fact that the Bernoulli sampling has been performed with the assumption that the probability of survival is the same for each period ($p \equiv 0.9$). This assumption, on the other hand, is equivalent to the assumption of agelessness, which, in turn, implies the exponential character of the lifetime distribution.

[8] Note that the expression 'spectrum point' is used here in the same sense as in Chapter 5. See in contrast footnote 1 after Eq. (1).
[9] See remark **(#14)** for the recipe.

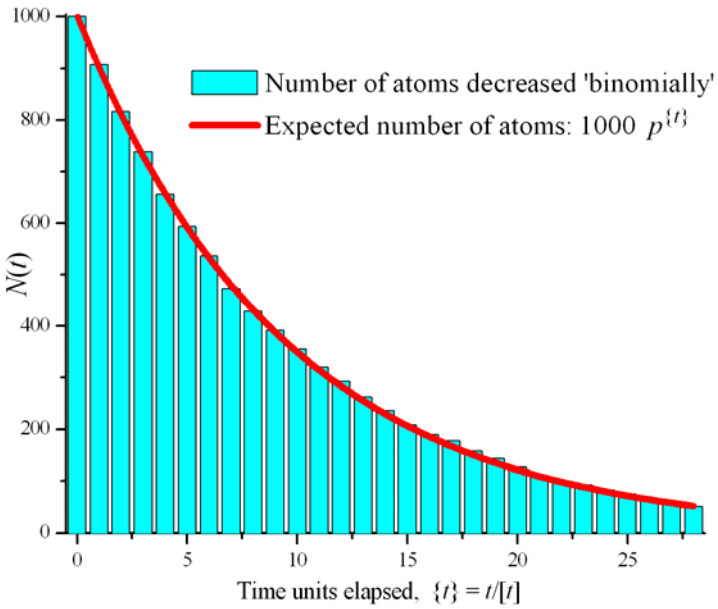

FIGURE 11. The possible fate of 1000 atoms followed by simulation. The probability that a given atom will survive a single time unit has been fixed at $p = 0.9$. The number of survivors decreases monotonically according to a step function that follows the curve of an exponential function. On the abscissa we use the standard notation recommended by IUPAP to express the physical quantity t as a product of a numerical value $\{t\}$ and a unit $[t]$. (See the text for more detail.)

The exponential function obtained for the expected value can be rewritten in e-based form as well with the notation $\{\lambda\} = \ln(1/p)$, and if we attribute the unit $[\lambda] = [t]^{-1}$ to the 'new' physical quantity λ (which we may call the 'decay constant') we can write:

$$\langle N\rangle(t) = N_0 p^{\{t\}} = N_0 \,\mathrm{e}^{-\ln(1/p)\{t\}} = N_0 \,\mathrm{e}^{-\{\lambda\}\{t\}} = N_0 \,\mathrm{e}^{-\lambda t}, \tag{82}$$

which can be considered as the stochastic reinterpretation of the decay law.

If the probability of decay ($q = 1-p$) is small (e.g. in the given case it is only 0.1), then the factor $\ln(1/p)$ can be expanded into a Taylor series, yielding a probabilistic interpretation for the decay constant λ:

$$\lambda[t] = \frac{\lambda}{[\lambda]} = \{\lambda\} = \ln\frac{1}{p} = -\ln(1-q) \approx q\,. \tag{83}$$

Thus **λ measures the probability that a selected atom will decay during the time unit**, provided that this probability (and, consequently, λ) is small enough ($q << 1$), which, in turn, depends on how long a time unit has been selected. In other words, the numerical value of the decay constant must be sufficiently small ($<< 1$) at the given choice of time unit, in order that the probabilistic interpretation can work. (Note that this interpretation corresponds to the second condition fixed at the interpretation of the exponential distribution.)

FIGURE 12 shows the change of the simulated $N(t)$ values, i.e. the simulated numbers of decay events occurring in the time unit. The corresponding expected value is:

$$\langle n\rangle(t) = -\Delta\langle N\rangle(t) = -N_0(p^{\{t\}} - p^{\{t\}-1}) = -N_0(p-1)p^{\{t\}-1} = N_0\frac{q}{p}p^{\{t\}} = N_0\frac{q}{p}\mathrm{e}^{-\lambda t}. \tag{84}$$

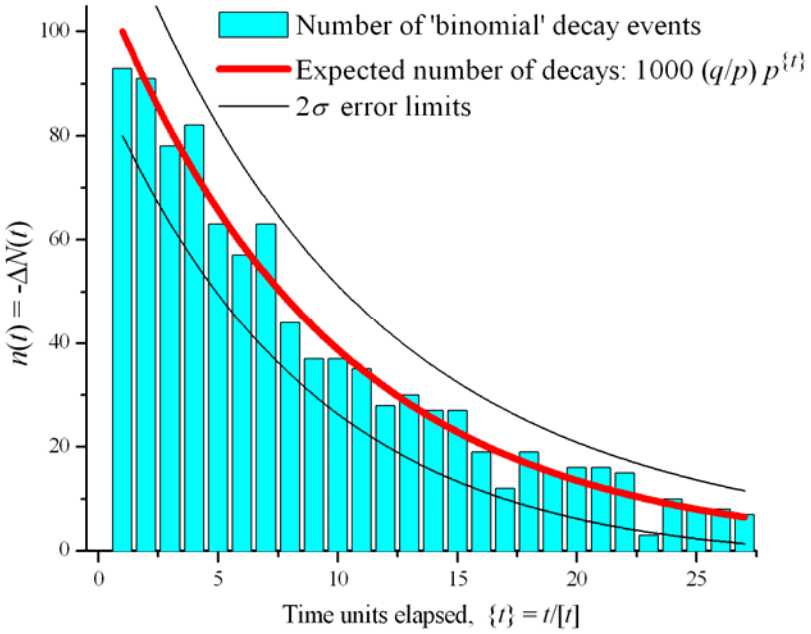

FIGURE 12. The numbers of decay events pictured have been obtained by forming the difference of the subsequent step-heights of the simulated step function in FIGURE 11. Note that the 'observed' numbers of decay events are not at all monotonically decreasing in time. However, fluctuating up and down, they follow the same exponential curve as the numbers of atoms themselves.

Comparing Eqs. (82) and (84) as well as FIGURE 11 and FIGURE 12 we can see that the time dependence of the number of decay events follows the same exponential rule as that of the number of intact atoms themselves.

3.4.3. The Poisson connection

If the parameter μ of the Poisson distribution means the expected number of pulses arriving over a certain time interval t, then the equation:

$$\nu \equiv \mu / t \tag{85}$$

will define the **mean frequency** of such pulses.

The probability that in the period t we count exactly x pulses arriving at a mean frequency ν can be calculated from the mass function of the $\Pi(\nu t)$ Poisson distribution:

$$P(X = x; t, \nu) = \frac{(\nu\, t)^x}{x!} \mathrm{e}^{-\nu t} \quad (x = 0, 1, 2, \ldots). \tag{86}$$

Hence the probability that no pulse is counted in the observation period t is:

$$P(X = 0; t, \nu) = \mathrm{e}^{-\nu t}. \tag{87}$$

Note that Eq. (87) also expresses the probability that the waiting time T (i.e. the time elapsed between subsequent pulses)—a continuous random variable—will be greater than t:

$$P(T > t) = \mathrm{e}^{-\nu t}. \tag{88}$$

Thus the distribution function of the waiting time T is identical with that of the $\gamma(1, \nu)$ exponential distribution:

$$F(t) = P(T \le t) = 1 - P(T > t) = 1 - \mathrm{e}^{-\nu t}. \tag{89}$$

(#36) This statement can be reversed, i.e. if the waiting time between pulses has a $\gamma(1,\nu)$ distribution, then the number $X(t)$ of the pulses counted over the period t is a random variable with a $\Pi(\nu t)$ Poisson distribution. Since the time is an explicit parameter here, we are dealing with a process. Accordingly, the 'function' $X(t)$ is referred to as **Poisson process** in the theory of stochastic processes.

(#37) Continuing remark **(#26)**: Consider a **radioactive specimen**. If the detector pulses (including the background) arrive at a mean frequency ν, then the **counts** registered over the counting period t have a $\boldsymbol{\Pi(\nu t)}$ **Poisson distribution**, and the **waiting times** between pulses have a $\boldsymbol{\gamma(1, \nu)}$ **exponential distribution**.

(#38) Continuing the previous remark: If the background can be neglected, then the mean frequency ν equals the source strength or observed activity. Thus it is the observed activity rather than the decay constant which directly appears in the exponential density function of the waiting times (see FIGURE 13). The value of the decay constant comes into the picture only indirectly through the expression:

$$\nu = \eta \lambda N \tag{90}$$

where η is the efficiency of the detecting system for the radiation produced by the decay and N is the number of radionuclei that are still intact.

(#39) The Poisson process can be used for the estimation of the count rate of **chance coincidences**. Let ν_1 and ν_2 denote the mean frequencies of pulses/signals coming from two independent detectors (signal channels). The **resolving time** of the **coincidence circuit** is θ, and the counting time is t. (A coincidence circuit only sends out an output signal, if—within the resolving time θ—an input signal coming from channel 1 is followed by an input signal from channel 2 or vice versa. Since the two types of successful events exclude each other, the coincidences estimated separately are to be added up in order to get the total number of coincidences over the counting time t.)

Suppose that an input signal is received from channel 1. The probability that no signal will arrive from channel 2 within the resolving time θ is $\exp(-\nu_2\theta)$, because the waiting times have exponential distribution. (It follows from the memorylessness of the exponential distribution that no matter when we start measuring the waiting time, we will get the same time distribution. In other words—see FIGURE 13—it does not matter whether we use the previous pulse of the very same sequence as a starting signal, or we just start measuring the time at random by leaving the business of starting to the pulses of another sequence.) The probability of the complementary event (i.e. that a signal will arrive from channel 2 within the resolving time θ) is $1 - \exp(-\nu_2\theta)$. Since the coincidence circuit is used for sorting out chance coincidences, we may assume that the exponent is so small that Taylor expansion is allowed. Hence the probability in question is $\nu_2\theta$. This probability can be considered as a type of '**efficiency of detection**' by which the coincidence circuit 'filters' (i.e. performs Bernoulli sampling on) the signals coming from channel 1. Since the latter have a $\Pi(\nu_1 t)$ Poisson distribution, the Bernoulli sampling will result in a number of chance coincidences with a $\Pi(\nu_1 t\, \nu_2\theta)$ distribution. Due to symmetry, we get the same result when channel 2 provides the start signal and channel 1 the stop signal. All in all, the total number of chance coincidences have a $\Pi(2\nu_1 t\, \nu_2\theta)$ distribution owing to the **addition theorem** of the Poisson

distribution. Thus the expected number of chance coincidences is about $2\theta\nu_1\nu_2\,t$, from which the **mean frequency of chance coincidences** is as follows:

$$\nu_{12} = 2\theta\nu_1\nu_2\,. \tag{91}$$

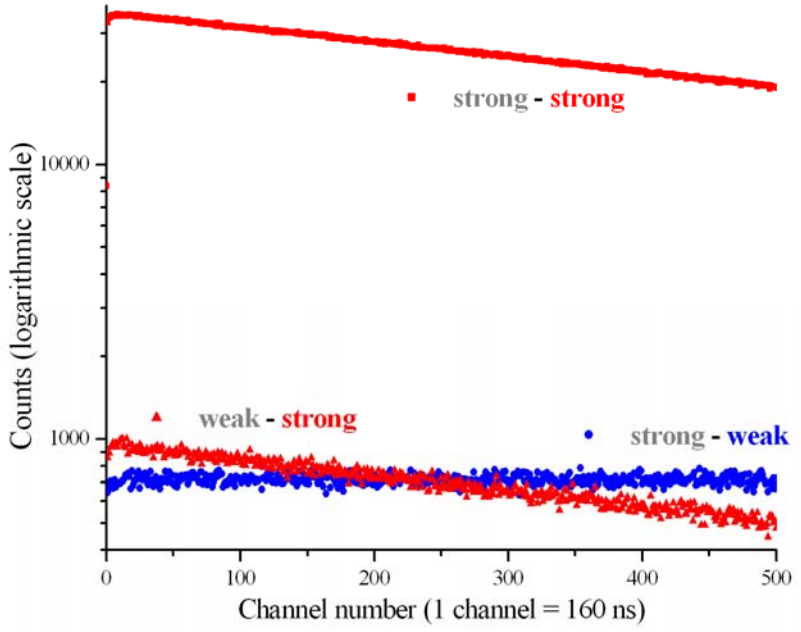

FIGURE 13. Experimental distribution of the **waiting times** between the pulses coming from two independent radioactive sources containing ^{22}Na. The labels (e.g. strong-strong) refer to the strengths (i.e. the observed activities) of the sources which produced the start signal and the stop signal, respectively, for the time measurement. We can see that the slope of each curve is determined by the mean frequency (strength) of the source producing the stop signals. Note also that the start signals are released at random moments and their sequence is completely independent of the sequence of the stop signals. This result well demonstrates the **memorylessness of exponential distribution**. The same experiment is usually suggested to produce 'white' noise (i.e. random signals that are uniformly distributed in time) for the PAS measurements. Our results clearly indicate, however, that the noise obtained this way can only be considered 'white' in the same way as a finite section of the exponential curve can be regarded horizontal.

(#40) It is a common mistake with 'beginners' that they directly apply the Poisson-type error estimation to the count rate $\nu = n/t$ calculated from the count n registered over a period of t in the belief that the 'error' of the count rate is $\pm\sqrt{\nu}$. Dimensional analysis, however, reveals that this cannot be the case, because if we choose to 'measure' ν in s^{-1}, then the unit of its 'error' calculated this way will be $\mathrm{s}^{-1/2}$, which does not make any sense. Actually, **the 'square-root' rule applies to 'unprocessed' counts** only, i.e. to n in the given case.

Now, the **correct error calculation of the count rate** ν goes like this. Let us use the notation recommended by IUPAP for scalar-type physical quantities, i.e., $a = \{a\}[a]$, where a is a physical quantity, $\{a\}$ is its numerical value and $[a]$ is the unit in which we happen to measure a (Cohen and Giacomo 1987). Thus the error of the counts is $\sqrt{n} = \sqrt{\nu\,t} = \sqrt{\{\nu\}[\nu]\,\{t\}[t]} = \sqrt{\{\nu\}\,\{t\}}$, because $[\nu] = [t]^{-1}$. Since ν is obtained from n by dividing it by t, the 'change-of-scale' rule of standard deviation yields for the error of ν $\sqrt{n}/t = \sqrt{\{\nu\}\,\{t\}}/(\{t\}[t]) = \sqrt{\{\nu\}/\{t\}}/[t]$. Note that the error calculation is performed on

dimensionless numbers with the time unit fixed for all the quantities involved. So the **correct data presentation** of count rates can be either $\left(\{\nu\} \pm \sqrt{\{\nu\}/\{t\}}\right)[t]^{-1}$ or simply $n/t \pm \sqrt{n}/t$.

Note also that, owing to the **erroneous practice** mentioned above (which, dimension aside, can be rewritten in the form $\{\nu\} \pm \sqrt{\{\nu\}}$ for better comparison), the standard deviation will be either over- or underestimated depending on whether the counting time t is shorter or, respectively, longer than the time unit used for the count-rate calculation. In the case of a one-minute measurement, e.g., if the time unit is the second, then we overestimate the error; if the unit is the hour, then we underestimate it; and we only get (numerically) correct result if the time unit happens to be the minute.

(#41) It is a fairly common task in nuclear science to tell **whether a hypothetical decay would occur**. The problematic part is represented by measurements, which have a negative outcome, when none of the atoms subjected to surveillance are found to decay during the observation period. Then the original question always remains open and new questions arise. (1) What would have happened, if the observation period had been longer [in which case Eq. (87) would have given more chance for a positive result]? (2) What would have happened, if we had subjected more 'suspicious' atoms to the surveillance (in which case Eq. (90) would have provided larger ν value, and therefore Eq. (87), again, would have increased the chances of a positive outcome)?

It is clear that the possibility of decay can never be excluded in cases like that. However, we can still specify a finite upper boundary ν_0, so that the 'real' frequency ν of the decay is lower than this limit with certain probability. Therefore this upper boundary depends on the **confidence level** α set by us (Leo 1987):

$$\nu_0 = \eta\lambda_0 N = -\frac{\ln(1-\alpha)}{t}. \tag{92}$$

For instance, if we want 90% certainty, then we are at the confidence level $\alpha = 0.9$.

Eq. (92) can be 'deduced' from Eq. (87) using common-sense considerations. Suppose that we want to give a guess at the value of ν. We only know that no signal could be detected over a certain observation period t. Then, the best thing we could do is form the weighted 'average' of all the possible ($0 \le \nu < \infty$) values of ν, using the weights provided by Eq. (87):

$$\overline{\nu} = \frac{\int_0^\infty \nu \mathrm{e}^{-\nu t}\,\mathrm{d}\nu}{\int_0^\infty \mathrm{e}^{-\nu t}\,\mathrm{d}\nu} = \int_0^\infty \nu t \mathrm{e}^{-\nu t}\,\mathrm{d}\nu$$

The advantage of this type of weighting is that it tends to suppress large values of ν, for which negative outcome is not likely to occur. On the other hand, it accentuates small values of ν, which are more compatible with the negative result.

Note that, according to Eq. (4), the integral on the right-hand side of the above equation has the shape of an expected-value formula with the (exponential) density function

$$f(\nu) = t\,\mathrm{e}^{-\nu t}$$

in which the variable and the parameter have swapped roles with each other. Hence the 'probability' of $\nu \le \nu_0$ is:

$$\alpha \equiv P(\nu \le \nu_0) = \int_0^{\nu_0} f(\nu)\,\mathrm{d}\nu = 1 - \mathrm{e}^{-\nu_0 t}$$

which yields Eq. (92).

3.5. Gamma distribution: $\gamma(r,\nu)$

The relation of gamma distribution to exponential distribution is similar to that of binomial distribution to Bernoulli distribution in so far as the former (i.e., gamma and binomial) is the convolution power of the latter (i.e., exponential and Bernoulli, respectively), while the latter is a special case of the former. Gamma distribution is also related to the Poisson process (see the interpretation given below). However, in the case of the Poisson process we are looking for the distribution of $X(t)$, the number of signals that arrive over a fixed period of time t. In the case of the gamma distribution, on the other hand, we are interested in the distribution of the time T that is needed for a given number of random signals to arrive.

Density function[10]: ..... $f_r(t) = \begin{cases} \dfrac{(\nu\, t)^{r-1}}{\Gamma(r)}\,\nu\,\mathrm{e}^{-\nu t} & \text{if} \quad t > 0 \\ 0 & \text{if} \quad t \le 0 \end{cases}$

Characteristic function: ..... $\varphi(u) = (1 - \mathrm{i}\,u/\nu)^{-r}$

Mean frequency: ..... $\tau_r^{-1} = \dfrac{\nu}{r}$

Expected value: .. $\tau_r \equiv \dfrac{r}{\nu}$ **Mode**: ..... $\dfrac{r-1}{\nu}$

Variance: ..... $\dfrac{r}{\nu^2}$ **Relative deviation**: ..... $\dfrac{1}{\sqrt{r}}$

Interpretation. Let T_1, T_2,… T_n. be a sequence of independent random variables with $\gamma(1,\ \nu)$ exponential distribution. Then the random variable defined by the sum $T = T_1 + \ldots + T_n$ has a gamma distribution of the nth order with parameter ν. In other words, T has a $\gamma(n,\ \nu)$ gamma distribution.

(#42) Note that, according to this interpretation, the exponential distribution of parameter ν can be rightly considered as a $\gamma(1,\nu)$ distribution, i.e. it belongs to the family of the gamma distributions (see FIGURE 14). Later on we will see that the sum of squares of independent random variables with $N(0,1)$ standard normal distribution also has a special type of gamma distribution called χ^2 distribution.

(#43) For the complete gamma function

$$\Gamma(r) \equiv \int_0^{\infty} \mathrm{e}^{-x}\, x^{r-1}\,\mathrm{d}\,x \tag{93}$$

we have the following **recursion formula**: $\Gamma(r+1) = r\,\Gamma(r)$.

In particular: $\Gamma(1/2) = \sqrt{\pi}$, $\Gamma(1) = 1$ and $\Gamma(n) = (n-1)!$ $(n = 1, 2, \ldots)$.

(#44) We can see that the relative deviation rapidly decreases with the increasing number of waiting periods ($n = r$). This property of the gamma distribution is utilized by the **scalers**,

[10] The order r and the parameter ν are positive real numbers, and $\Gamma(r)$ is the complete gamma function.

which only produce an output signal on the arrival of the *n*th input signal. (In this context *n* is called the **scaling factor**.) The actual signal frequencies of scalers are, therefore, much more uniform (i.e. the signals follow the 'rhythm' of the mean frequency much more 'smoothly'), than the original signals themselves (see FIGURE 14).

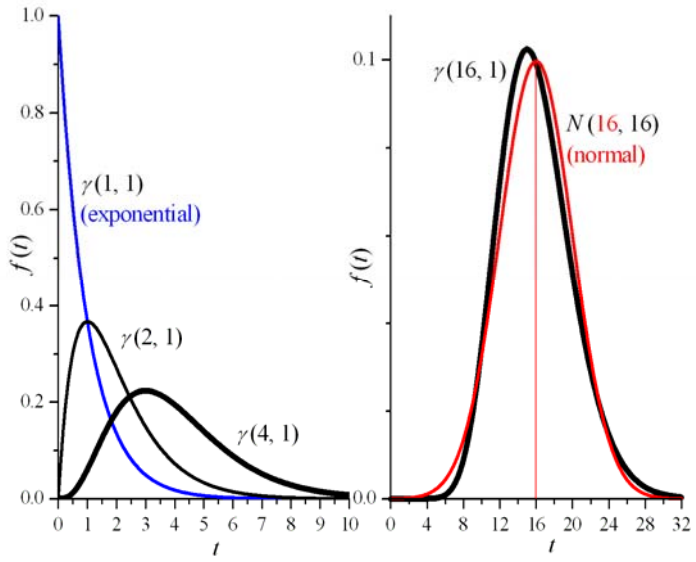

FIGURE 14. Density functions of **gamma distributions** of integral order with the parameter value fixed at $\nu = 1$. The order $r = 1$ is characterized by the well-known asymmetric density function of the **exponential distribution**. However, as the order increases, the asymmetry decreases, and for as low an order as $r = 16$ we get an acceptable fit with an appropriately chosen **normal distribution**. The gamma distribution is characteristic of the waiting times between **scaled events** (scaler signals). In that context the order r is called the **scaling factor**, and ν is the **mean frequency** of the original (i.e. unscaled) signals.

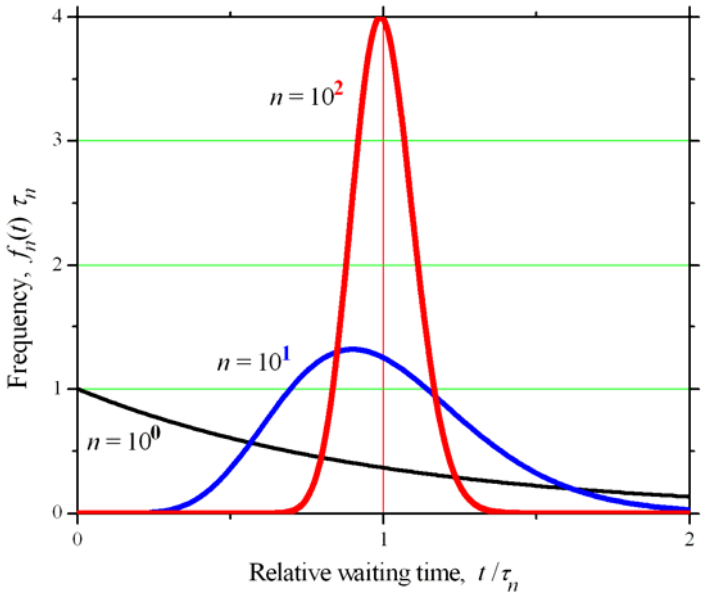

FIGURE 15. When counting the particles/photons emitted by a radioactive specimen, each decimal digit of the counter switches to the next figure after a **waiting time** determined by a **gamma distribution** the order of which (n), in turn, is determined by the decimal place of the digit. Note that a narrower

distribution means a steadier rhythm of switching. For better comparison, the density functions have been renormalized so that the geometric area under each curve is the same.

(#45) The **decimal display** of regular **counters** is in fact a realization of a series of scalers with scaling factors $n = 10^0, 10^1, 10^2$, etc. (see FIGURE 15). Consequently, while the 'switching times' of the lowest digit ($n = 10^0 = 1$) have exponential distribution in accordance with the Poisson process, those of the rest are characterized by the density functions of the gamma distributions of the order $n = 10^1, 10^2$, etc.:

$$f_n(t) = \frac{(\nu t)^{n-1}}{(n-1)!}\,\nu\,\mathrm{e}^{-\nu t}\,. \tag{94}$$

Thus the relative deviation of the 'switching time' of the third (10^2) digit is about 10%, and that of the fifth (10^4) digit is only about 1%, which means that the figures at the fifth decimal place seem to 'rotate' almost steadily.

3.5.1. Properties

Addition theorem. If X_1 and X_2 are independent random variables with $\gamma(r_1, \nu)$, and $\gamma(r_2, \nu)$ gamma distribution, respectively, then the random variable $X = X_1 + X_2$ has a $\gamma(r_1 + r_2, \nu)$ gamma distribution.

Normal approximation. For large enough values of r, the $\gamma(r, \nu)$ gamma distribution can be approximated by an $N(r/\nu, r/\nu^2)$ normal distribution (see the graph in FIGURE 14 to the right). (If the order $r = n$ is an integer, then this follows directly from the interpretation and from the central limit theorem.) The normal probabilities for deviations from the mean ($\mu \pm d\sigma$) can be found in Table 3.

3.6. Normal distribution: $N(\mu, \sigma^2)$

Density function[11]: ..... $f(x) = \frac{1}{\sigma} f_{01}\left(\frac{x-\mu}{\sigma}\right) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]$

Distribution function: .. $F(x) = F_{01}\left(\frac{x-\mu}{\sigma}\right) = \frac{1}{2}\left[1 + \operatorname{erf}\left(\frac{1}{\sqrt{2}}\frac{x-\mu}{\sigma}\right)\right]$

Error function: ..... $\operatorname{erf} x = -\operatorname{erf}(-x) = \frac{2}{\sqrt{\pi}} \int_0^x \mathrm{e}^{-t^2}\, \mathrm{d}t$

Characteristic function: $\varphi(u) = \exp\left(\mathrm{i}\mu u - \frac{\sigma^2}{2}u^2\right)$

Expected value, **median**, **mode**, **center of symmetry**: μ

Variance: ..... σ^2 **Mean abs. deviation**: $\sqrt{2/\pi}\,\sigma \approx 0.798\sigma$

Interquartile range: 1.348σ ..... **Point of inflection of *f***: $\mu \pm \sigma$

Full Width at Half Maximum (FWHM):..... $2\sqrt{2\ln 2}\sigma \approx 2.355\sigma$

Interpretation. Let X be any random variable with existing expected value (μ) and variance (σ^2). Let X_i denote the random variable representing the outcome of the ith (i = 1, 2, … n) independent experiment to observe the value of X. (The random variables X_i are obviously independent and have the same distribution as X.) Then the sample mean $\overline{X} \equiv (X_1+X_2+\ldots+X_n)/n$ is asymptotically normal with $N(\mu, \sigma^2/n)$ distribution[12].

(#46) The above interpretation is the Lindberg-Lévy version of the **central limit theorem**. We could also have interpreted the [$N(np, npq)$] normal distribution as the limiting case of the $B(n,p)$ binomial distribution (DeMoivre-Laplace theorem), but that is only a special case of the general theorem phrased for sums when the limiting distribution in general is $N(n\mu, n\sigma^2)$ normal.

(#47) When the normal distribution is referred to as the limiting case of an integral valued (x = 0, 1, 2, …) discrete distribution (e.g. Poisson or binomial distribution), then the continuous density function $f(x)$ is actually a substitute for the **mass function** p_x. This is so, because the latter can be considered as a step function jumping at integral values. Therefore, since $\Delta x = 1$, we can write $p_x \approx f(x)\Delta x \equiv f(x)$ whenever x is an integer.

3.6.1. Properties

Standardization. Shifting and/or rescaling a normal random variable will result in another normally distributed random variable. Therefore, the standardization of the $N(\mu, \sigma^2)$ normal random variable X will result in an $N(0, 1)$ **standard normal random variable** Y:

$$Y = \frac{X-\mu}{\sigma}. \tag{95}$$

[11] f_{01} and F_{01} are the $N(0, 1)$ standard normal density function and distribution function, respectively.
[12] In physics and engineering etc. normal distribution is often called **Gaussian distribution**. However, we will only use the expression **Gaussian curve** meaning an un-normalized normal distribution function as shown in FIGURE 18.

(#48) The random numbers produced by common calculators usually have a uniform distribution between 0 and 1. However, such $U(0, 1)$ random numbers can be easily converted to **normally distributed random numbers** by summing them n at a time [see remark **(#46)**]. Considering large enough values for n, the summation will result in random numbers with an $N(n/2, n/12)$ distribution [see remark **(#8)**]. Thus, for $n = 12$ we get an $N(6, 1)$ distribution. In this case we can convert the result to $N(0, 1)$ **standard normal distribution** by subtracting 6 from the sum.

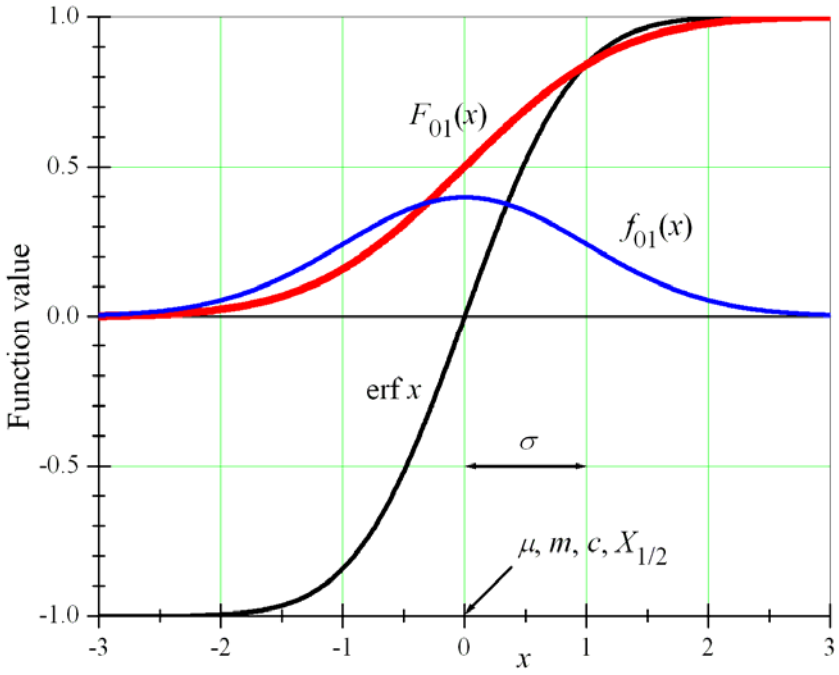

FIGURE 16. Comparison of the **error function** (erf) with the **density function** (f_{01}) and the **distribution function** (F_{01}) of the $N(0, 1)$ **standard normal distribution**. The intervals 0 ± 1, 0 ± 2, and 0 ± 3 correspond to the error limits $\mu \pm \sigma$, $\mu \pm 2\sigma$, and $\mu \pm 3\sigma$, respectively. The normal probabilities for deviations from the mean ($\mu \pm d\sigma$) can be found in Table 3.

(#49) The $N(0, 1)$ random numbers simulated according to the above remark can be used, e.g., for the **simulation of nuclear spectra**. Nuclear spectrum points (see in Chapter 5) have $N(\mu_i, \mu_i)$ distribution. If the μ_i values are calculated from the fitting function and R_i is an $N(0, 1)$ random number, then the formula $Y_i \equiv \mu_i + \sqrt{\mu_i} R_i$ defines a random number with $N(\mu_i, \mu_i)$ distribution as required.

Addition theorem. If X_1 and X_2 are independent random variables with $N(\mu_1, \sigma_1^2)$, and $N(\mu_2, \sigma_2^2)$ normal distribution, respectively, then the random variable $X \equiv X_1 + X_2$ has an $N(\mu_1+\mu_2, \sigma_1^2+\sigma_2^2)$ normal distribution.

Bernoulli sampling. Suppose that the aim of an experiment is to observe certain objects (e.g. all the 'visible' photons belonging to a light pulse that was produced as a result of the detection of a single radiation particle in the scintillator). Suppose that the number of these objects is a random variable with $N(\mu, \sigma^2)$ distribution. The objects are taken one by one and either 'accepted' with probability η or 'discarded' with probability $1-\eta$. In other words: a $B(1, \eta)$ Bernoulli sampling is applied to them. (Returning to the above example, η can represent the probability that a 'visible' photon will produce a photoelectron due to its interaction with the photocathode of the multiplier.) The question is, how many objects have been accepted altogether in the experiment, i.e. how many objects have passed the test of

Bernoulli sampling (e.g. how many photoelectrons have formed on the photocathode as a result of the detection of a single radiation particle) and how these accepted numbers are distributed if a whole series of like experiments is considered. The $N(\mu, \sigma^2)$ distribution mentioned above is, of course, the integral-valued 'discrete' version of normal distribution, for which the substitution values $f(n)$ of the density function act as mass function at the integral values of the x axis. Therefore the question can be answered by reversing the DeMoivre-Laplace theorem, i.e., if $\mu > \sigma^2$ and $\mu >> 1$, then there is a $B(n,p)$ binomial distribution which provides a fair approximation for the $N(\mu, \sigma^2)$ distribution. In the case of the binomial distribution, we have seen that the result of Bernoulli sampling is a $B(n, p\eta)$ distribution. Thus we only need to find the $N(\mu_\eta, \sigma_\eta^2)$ distribution that approximates this Bernoulli-sampled binomial distribution according to the DeMoivre-Laplace theorem. It is easy to show that the parameters of the 'sampled' and 'un-sampled' normal distributions are related as follows:

$$\mu_\eta = \eta\mu\,; \qquad \sigma_\eta^2 = \eta^2\sigma^2 + \eta(1-\eta)\mu\,; \qquad \frac{\sigma_\eta}{\mu_\eta} = \frac{\sigma}{\mu}\sqrt{1+\frac{1-\eta}{\eta}\frac{\mu}{\sigma^2}}\,. \tag{96}$$

(#50) According to the above formula, the normal distribution can broaden considerably as a result of Bernoulli sampling. Referring to the above example, this means that the **resolution of scintillation detectors** is limited (among others) by the fact that typically only 1 'visible' photon out of 10 produces a photoelectron at the photocathode. In the case of the NaI(Tl) scintillator, moreover, there is also a previous Bernoulli sampling, i.e. only one excitation out of three will produce a photon that is suitable for the production of photoelectrons at all. The fact that the number of excitations (occurring, e.g., due to the absorption of a gamma photon) has a normal distribution will be shown in the frame of the **renewal process**.

3.7. The χ^2 distribution: $\chi^2(k)$

Density function: ..... $f_k(x) = \begin{cases} \dfrac{x^{k/2-1}\,\mathrm{e}^{-x/2}}{2^{k/2}\,\Gamma(k/2)} & \text{ha} \quad x > 0 \\ 0 & \text{ha} \quad x \le 0 \end{cases}$

Characteristic function: ..... $\varphi(u) = (1-2\mathrm{i}u)^{-k/2}$

Expected value: .....k **Mode**:..... $k-2$

Variance: .....$2k$ **Relative deviation**:..... $\sqrt{\dfrac{2}{k}}$

Interpretation. Let Z_1, Z_2, … Z_k be a sequence of independent $N(0, 1)$ random variables. Then the sum of squares $Z_1^2+Z_2^2+\ldots+Z_k^2$ produces a random variable having a χ^2 distribution with k **degrees of freedom**—in short: $\chi^2(k)$ distribution.

(#51) If Y_1, Y_2, … Y_k are a sequence of independent random variables having the same $N(\mu, \sigma^2)$ distribution with exactly known expected value and variance, then the sum of squares of the standardized variables will have a $\chi^2(k)$ distribution.

(#52) Continuing the previous remark: If μ is not known, and therefore it is replaced by the average $\overline{Y} = (Y_1+Y_2+\ldots+Y_k)/k$ when forming the standardized variables, then the sum of squares of the standardized variables will have a $\chi^2(k-1)$ distribution.

(#53) If Y_1, Y_2, … Y_k are a sequence of independent random variables which have $N(\mu_i, \sigma_i^2)$ normal distributions ($i = 1, 2, \ldots, k$) with known expected values and variances, then the sum of squares of the standardized variables will have a $\chi^2(k)$ distribution.

(#54) Continuing the previous remark: Suppose that there is a model function $\mu(x) = \mu(x; b_1, b_2, \ldots b_n)$ which is linear in all of its n parameters ($b_1, b_2, \ldots b_n$) and that model function (now: fitting function) produces the expected values μ_i ($i = 1, 2, \ldots k$) whenever the parameter vector $\boldsymbol{b}$ is set to the value $\boldsymbol{b} = \boldsymbol{a}$:

$$\mu_i = \mu(x_i; a_1, a_2, \ldots a_n)\,. \tag{97}$$

Let us consider now the sum of squares

$$S_k = \sum_{i=1}^{k}\left(\frac{Y_i - \mu(x_i; b_1, b_2, \ldots b_n)}{\sigma_i}\right)^2\,. \tag{98}$$

Suppose that the parameter values $\hat{a}_1, \hat{a}_2, \ldots \hat{a}_n$ minimize the above sum. Then the minimized sum of squares

$$\hat{S}_k = \sum_{i=1}^{k}\left(\frac{Y_i - \mu(x_i; \hat{a}_1, \hat{a}_2, \ldots \hat{a}_n)}{\sigma_i}\right)^2 \tag{99}$$

has a $\chi^2(k-n)$ distribution (Press *et al.* 1999).

(#55) Continuing the previous remark: According to remark **(#53)** the sum of squares taken at the parameter values a_1, a_2, … a_n would have a $\chi^2(k)$ distribution because in this case the 'exact' expected values (μ_i) would show up in the sum. We stress this point because we will see at the evaluation of nuclear spectra that the values $\hat{a}_1, \hat{a}_2, \ldots \hat{a}_n$ are the **maximum**

likelihood estimates of the 'exact' values $a_1, a_2, \ldots a_n$. Therefore one might expect that the relation of $\hat{\boldsymbol{a}}$ to the concrete measured spectrum is the same as that of the parameter vector **a** which it estimates. Well, the decrease of the degree of freedom indicates that this assumption is false. The reason is that minimization tends to divert the estimated values from the 'exact' parameter values so that they can attribute the largest possible likelihood to the concrete spectrum. And this will be so even if the concrete spectrum has a rather low likelihood when calculated with the 'exact' values of the parameters.

(#56) The symbol $\Gamma(k/2)$ showing up in the density function denotes the complete gamma function. This is not by chance because the χ^2 distribution is actually a special case of the gamma distribution. (Looking at the density function f_k, we can easily recognize it as the density function of the $\gamma(k/2, 1/2)$ gamma distribution. The characteristic functions are still easier to compare.)

(#57) Note that when the degree of freedom (k) is large, then the expected value is practically equal to the mode, i.e. the expected value doubles as the 'most probable'[13] value. When evaluating nuclear spectra, χ^2 distributions with about 2000 degrees of freedom quite often occur (2048-channel spectra). In such cases the relative deviation is about 3%, i.e. the value of the $\chi^2(2000)$ random variable is most probably $2000 \pm 3\%$ (2000 ± 60). For more detail see the next remark. (The p-quantiles for χ^2 distributions $\chi^2(k)$ with degree of freedom $1 \le k \le 30$ can be found in Table 4.)

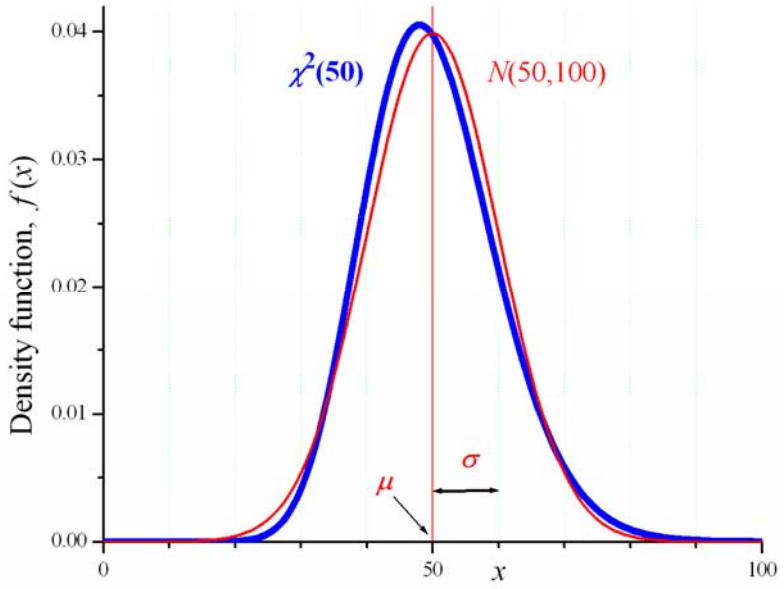

FIGURE 17. The $\chi^2(k)$ distribution can be fairly well approximated with the $N(k, 2k)$ distribution for as low a degree of freedom as $k = 50$.

3.7.1. Properties

Normal approximation. It follows from the interpretation as well as from the central limit theorem that for large enough values of k ($k \ge 30$) the $\chi^2(k)$ distribution can be approximated with the $N(k, 2k)$ normal distribution (see FIGURE 17).

[13] The reason for the quotation marks is that with continuous distributions any x value has 0 probability. What this statement really means is that the probability $dP = f(x)dx$ is at its maximum (dx is constant).

(#58) Continuing the previous remark: According to the properties of the normal distribution (see Table 3), the value of the $\chi^2(2000)$ random variable will be $2000 \pm 3\%$ (2000 ± 60) in 68% of the cases, $2000 \pm 6\%$ (2000 ± 120) in 95% of the cases, and $2000 \pm 9\%$ (2000 ± 180) in 99.7% of the cases (i.e. nearly always).

(#59) The **relative χ^2** is the scaled-down version of the $\chi^2(k)$ random variable with the scaling factor $1/k$. Consequently, both the expected value and the standard deviation decrease by the same factor $1/k$ resulting in $\mu_{\text{rel}} = 1$ and $\sigma_{\text{rel}} = \sqrt{2/k}$, respectively. Continuing the previous remark: this means that the value of the relative χ^2 will be 1 ± 0.03 in 68% of the cases, 1 ± 0.06 in 95% of the cases, and 1 ± 0.09 in 99.7% of the cases (i.e. nearly always). Relative χ^2 values out of this range are to be considered therefore with suspicion. (Of course, if k is smaller, then the error range is broader.)

Addition theorem. If X_1 and X_2 are independent random variables having χ^2 distributions with k_1 and k_2 degrees of freedom, respectively, then the sum $X = X_1 + X_2$ has a χ^2 distribution with $k_1 + k_2$ degrees of freedom.

3.8. Cauchy distribution: $C(m,\gamma)$

Density function: ..... $f(x)=\dfrac{1}{\pi\gamma}\dfrac{1}{1+\left(\dfrac{x-m}{\gamma}\right)^2}$

Distribution function: ..... $F(x)=\dfrac{1}{2}+\dfrac{1}{\pi}\operatorname{arctg}\dfrac{x-m}{\gamma}$

Characteristic function: ..... $\varphi(u)=\exp\left(\mathrm{i}mu-\gamma|u|\right)$

Expected value, **variance**, **standard deviation**: .....☹

Median, **mode**, **center of symmetry**: .....m

Full width at half maximum (FWHM), **interquartile range**: 2γ

Point of inflection of the density function: ..... $m\pm\gamma/\sqrt{3}$

All of the distributions mentioned so far lie within the range of attraction of normal distribution, i.e. the central limit theorem is valid for distributions that are quite dissimilar to the normal distribution such as the discrete Bernoulli distribution and the asymmetric exponential distribution. The Cauchy distribution presented below is, on the contrary, very similar in shape to the normal distribution (see FIGURE 18), but it has neither an expected value, nor a (finite) variance, and therefore is exempt from the rule.

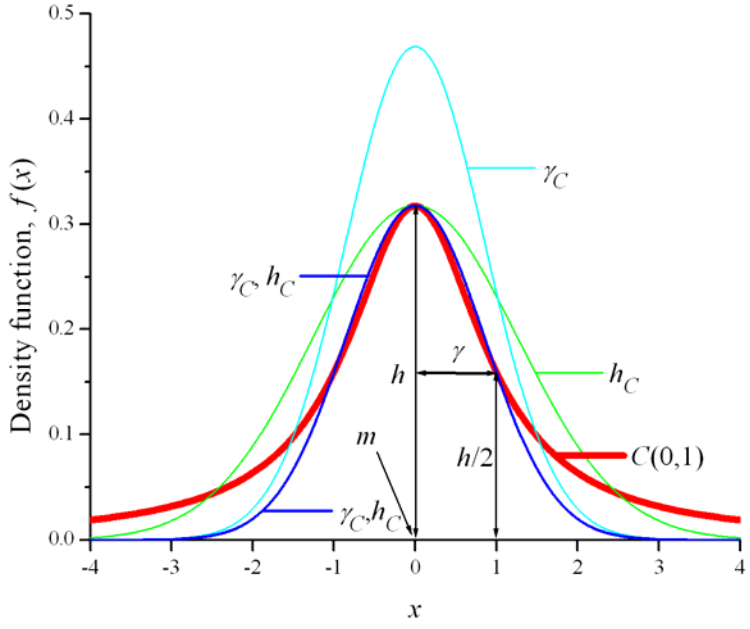

FIGURE 18. The **density function of the *C*(0, 1) Cauchy distribution** (thick red line, also called Lorentzian curve) compared with different Gaussians. The green and turquoise curves indicate **normal density functions** (the area under each curve is 1). In one case the height (h_C), in the other case the width (γ_C) is fixed to be equal with that of the Lorentzian. The thicker blue curve shows a Gaussian having the same height and width as the Lorentzian (γ_C, h_C). (With both parameters fixed, the Gaussian could not be normalized. Thus the area under the curve is 0.678 this time.) The difference between the distributions is quite obvious here: the **Lorentzian** approaches 0 much more slowly than the **Gaussian**, which is in accordance with the nonexistence of the expected value and the variance in the case of the Cauchy distribution.

Interpretation. Let Θ be a random variable uniformly distributed in the interval $(-\pi/2, \pi/2)$. Then the random variable

$$Y = m + \gamma \tan\Theta \tag{100}$$

has a $C(m, \gamma)$ Cauchy distribution. In other words: let us consider the xy-plane and two lines—one (line 1) passing through the origin and another (line 2) parallel to the y-axis and intersecting the x-axis at the point $x = 1$. Let us spin line 1 around the origin at constant angular velocity. Let us measure the y-coordinate of the intersection of lines 1 and 2 at random moments. The Y values obtained in this way will have a $C(0, 1)$ distribution (Hogg and Tanis 1988).

(#60) Using the above interpretation, it is easy to convert uniformly distributed $U(-\pi/2, \pi/2)$ random numbers to **random numbers with Cauchy distribution**. Such a simulated sequence (200 data) is shown in FIGURE 19, together with as many normally distributed random numbers having the same FWHM. Note that all of the normal random numbers lie within a few FWHMs from the origin. The Cauchy-type random numbers, on the other hand, behave in a much more disorderly way, i.e. there are quite a number of points that are way out of the same range.

(#61) Another possibility for interpretation (and simulation) is provided by the formula $X = Y_0/|Y_1|$, where Y_0 and Y_1 are independent random numbers/variables with $N(0, 1)$ distribution. This transformation results in $C(0, 1)$ distribution. The recipe is based on the fact that the Cauchy distribution can be considered as a special case of the (Student's) ***t*** **distribution** (not discussed here).

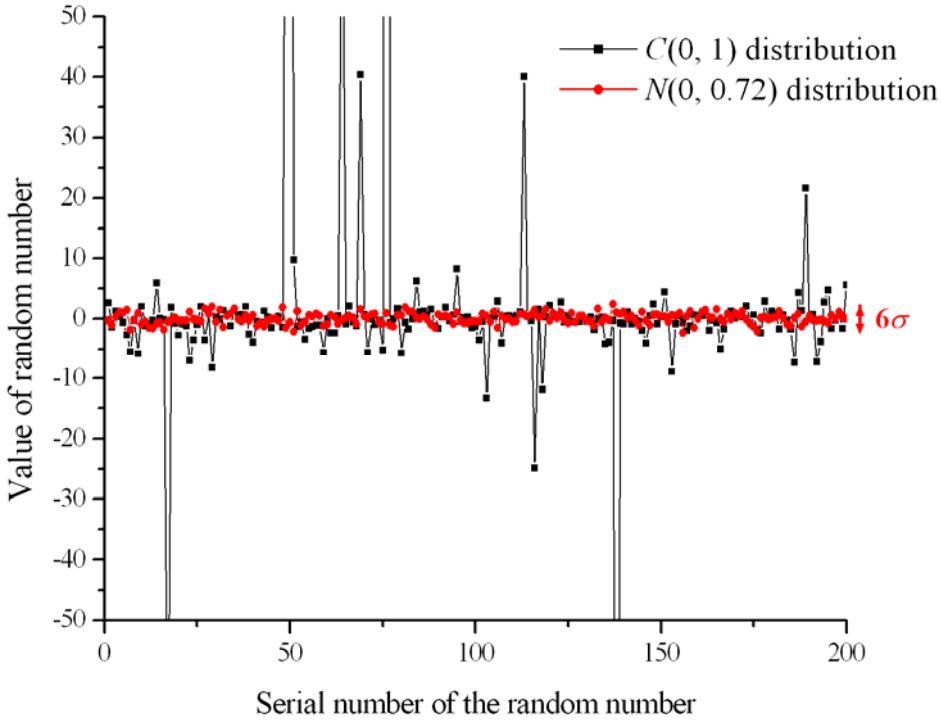

FIGURE 19. Comparison of normal random numbers with random numbers having Cauchy distribution. Both distributions have the same halfwidth (FWHM = 2). Note that some of the Cauchy-type random numbers are as large as 500 in this particular sequence. On the other hand, all of the normal random numbers lie in the narrow range of $6\sigma = 5$ determined by the $0 \pm 3\sigma$ limits.

(#62) In Mössbauer spectroscopy, the density function of Cauchy distribution is called a **Lorentzian curve**[14]. This curve is characteristic of the energy uncertainty of excited (nuclear) states, which follows from the fact that excited states have exponential distribution with a finite mean life τ. The **natural linewidth** Γ, i.e. the FWHM of the Lorentzian energy density, is twice of the parameter γ ($\Gamma = 2\gamma$).

(#63) The **lack of expected value and variance** can be quite a shock for those who had been familiar with the physical meaning of the Lorentzian curve but learned about this mathematical 'defectiveness' at a later date.[15] This is so, because the nonexistence of these parameters makes them believe that the distribution is hopelessly smeared over the energy axis. However, this is not the case, which is quite clear from the viewpoint of physics, because otherwise no sharp spectrum lines could exist. The expected value (and the variance built on that) fails, because the 'central tendency' of the distribution is simply not sufficient for the convergence of the respective formula(e)[16]. Other parameters describing the 'location' and 'dispersion' of the distribution (mode, interquartile range, etc.) still may and actually *do* work.

(#64) Continuing remarks **(#62)-(#63)**: The formula $\tau\Gamma = \hbar$ is often used for the estimation of the **natural linewidth**. This formula is sometimes interpreted as the time-energy equivalent of the **Heisenberg relation**, where τ is the uncertainty (standard deviation) of the lifetime and Γ (FWHM) is that of the energy state. We should point out, however, that while τ can play the assigned role (because the standard deviation is equal to the expected value in the case of the exponential distribution), the quantity Γ **cannot be interpreted as standard deviation**, since the Cauchy distribution does not have any.

Nevertheless the formula works not only in theory but also in practice as is proven for 30 or so nuclides for which the values of Γ and τ have been determined by independent methods (Belgya *et al.* 1993).

3.8.1. Properties

Standardization. The rescaling and shifting of Cauchy distribution results in another Cauchy distribution. Therefore the following transformation of the random variable X having a $C(m, \gamma)$ Cauchy distribution

$$Y = \frac{X - m}{\gamma} \tag{101}$$

generates a random variable Y with $C(0, 1)$ Cauchy distribution.

(#65) Note that in the above formula, the 'substitutes' of the expected value and standard deviation appear as location and dispersion parameters.

Addition theorem. If X_1 and X_2 are independent random variables with $C(m_1, \gamma_1)$ and $C(m_2, \gamma_2)$ Cauchy distribution, respectively, then the random variable $X = X_1 + X_2$ has a $C(m_1+m_2, \gamma_1+\gamma_2)$ Cauchy distribution.

(#66) In **Mössbauer spectroscopy**, the peak-shape of the **transmission spectrum** (measured with a thin absorber, see FIGURE 22) is described as the **convolution of the Lorentzians** characteristic of the source and the absorber. According to the addition theorem,

[14] In particle physics, the same function is called the Breit–Wigner curve (Lyons 1986).
[15] The confrontation with this defectiveness often fails to occur. For instance, Bevington, in his often-cited work (Bevington 1969), writes about Cauchy distribution as if it had an expected value.
[16] Note that the density function is asymptotically proportional to x^{-2}, and therefore the integrand in the expected value formula is proportional to x^{-1}, hence the integral itself ($\sim \ln x$) is boundless.

the result of such a convolution will be another Lorentzian with a halfwidth equal to the sum of both halfwidths. In other words, in this case, the FWHM is twice of the natural linewidth Γ.

(#67) Note that in the case of the Cauchy distribution, the scaling-up of the random variable with the factor 2 leads to the same result as the addition of two independent random variables with the same distribution (i.e. both parameters m and γ will double). In the case of distributions 'more normal' than the Cauchy distribution, on the other hand, the width will only increase by the factor $\sqrt{2}$ when two like (although independent) variables are added up. As shown before, this is due to the 'Pythagorean theorem' of the standard deviations, by which we mean that it is the variances that add up rather than the standard deviations themselves.

4. Applications of Stochastic Processes

In this chapter we will give a few examples to show the applicability of stochastic processes in nuclear science. We should stress that our goal is to present *illustrations* rather than to derive formulae that serve as the final solutions for the problems cited. We hope that reading the next couple of pages can help to arouse interest in this field.

4.1. Renewal processes

The 'prototype' of the renewal process is the following. Consider a part of equipment the lifetime of which, T, is a random variable with expected value μ and variance σ^2. The first such part started to work at time 0. Having broken down at time T_1, it was promptly replaced with a new part, which was working for a period T_2, i.e. it broke down at time T_1+T_2, and so on. The nth renewal took place at time: $S_n = T_1+\ldots+T_n$. The **renewal process** N_t counts the number of renewals occurring in the time interval $[0, t]$, i.e.:

$$N_t = n, \text{ if } S_n \le t < S_{n+1}, \text{ where } n = 0, 1, 2, \ldots \tag{102}$$

We cite two important theorems concerning the above process:

Renewal theorem. $E(N_t) \approx t/\mu$, if t is large enough. (This is in agreement with the expectation based on common sense: If a part works for a period μ on an average, then we will need about t/μ parts over the total operating period t.)

Central limit theorem for renewals. N_t is asymptotically normal with $N(t/\mu, t\sigma^2/\mu^3)$ distribution.

4.1.1. The Poisson process as renewal

It is not difficult to recognize the renewal process in the Poisson process discussed earlier, where the random variable T had a $\gamma(1, \nu)$ exponential distribution, as a consequence of which N_t (or, with the earlier notation: X) turned out to have a $\Pi(\nu t)$ Poisson distribution. The Poisson process represents one of the few cases where the distribution of the renewal process is known for any value of t. The importance of the **central limit theorem for renewals** lies in the very fact that, for large enough values of t, we can count on normal distribution even if we do not know the exact distribution of N_t. The validity of the theorem can be easily checked in this particular case, since by substitution we get the well-known result that for large enough μ we can use the approximation $\Pi(\mu) \approx N(\mu, \mu)$.

(#68) The fact that the Poisson process can be interpreted as a renewal makes it possible to generate **random numbers with Poisson distribution** using exponentially distributed random numbers (Goodman 1988, Lux and Koblinger 1991).[17] If, namely, T_1, T_2, … are a sequence of random numbers/variables with a $\gamma(1, \nu)$ exponential distribution, then the random number/variable N determined by the condition

$$T_1 + T_2 + \ldots + T_N \le t < T_1 + T_2 + \ldots + T_N + T_{N+1} \tag{103}$$

will have a $\Pi(\nu t)$ Poisson distribution according to inequality (102).

[17] For the generation of exponentially distributed random numbers see remark (#30).

(#69) The discrete equivalent of the Poisson process is related to a **series of Bernoulli trials**, in which case the individual trials (e.g., coin tosses) can be assigned to individual discrete moments (i.e. to the serial number of the toss, n). It is clear that the process $X(n)$ has a $B(n, p)$ binomial distribution—the distribution characteristic of the number of heads turning up in a series of n tosses.

4.1.2. Consideration of dead time

The Poisson process is connected with a (non-existent) type of particle/signal detection, when the detecting system is promptly able to deal with a new signal after having received the previous one. In reality, however, any detecting system can only deal with a new signal after a certain period of time Θ_i called the **dead time**. During this period (which should be treated as a random variable in general), the detecting system is recovering (e.g. GM counter), or it is busy with signal processing (e.g. pulse height analysis). The dead time is followed by the **waiting time** T_i having exponential distribution, the same as the one discussed at the Poisson process (see Eq. (89)), with the same mean frequency ν. (We should again refer to the memorylessness of the exponential distribution.) The total waiting time (including the dead time) connected with the ith (actually detected) signal therefore is $Z_i = \Theta_i + T_i$. Thus the nth signal is detected at the moment $S_n = Z_1 + \ldots + Z_n$.

Considering the independence of Θ_i and T_i, as well as that the distributions of both Θ_i and T_i are concentrated to the half-line $t > 0$, the distribution function of Z_i is given by the following convolution (see Eq. (51)):

$$F(z) = G * h(z) = \int_0^z G(z-x)h(x)\,\mathrm{d}x = \int_0^z G(z-x)\nu\,\mathrm{e}^{-\nu x}\,\mathrm{d}x$$

where G is the distribution function of Θ_i, and h is the (exponential) density function of T_i. The subscripts have been omitted, because the Θ_is have the same distribution and so do the T_is (separately). The density function of $Z_i\,(f)$ can be obtained by differentiation. The density function of $S_n\,(f_n)$ can be obtained by forming the nth convolution power of $f = F'$:

$$f_n(z) = f * \ldots * f(z)$$

In particular, if the dead time is constant (θ), then $f = \nu \exp[-\nu(t-\theta)]$, which is the exponential density function shifted to θ ($t > \theta$). Therefore S_n has a shifted gamma distribution with the following density function:

$$f_n(t) = \frac{[\nu\,(t-n\theta)]^{n-1}}{(n-1)!}\,\nu\,\mathrm{e}^{-\nu(t-n\theta)} \quad (t > n\theta)\,.$$

The following results hold even if the concrete distribution of the process N_t is unknown:

$$\mu \equiv E(Z) = E(\Theta) + E(T) = \theta + \frac{1}{\nu} \tag{104}$$

$$\sigma^2 \equiv D^2(Z) = D^2(\Theta) + D^2(T) = D^2(\Theta) + \frac{1}{\nu^2} \tag{105}$$

where θ is the **mean dead time**, and ν is the **mean signal frequency** (mean count rate with zero dead time). If the dead time is constant, the latter equation becomes simpler:

$$\sigma^2 = \frac{1}{\nu^2}. \tag{106}$$

The mean count rate in the case of constant dead time θ is:

$$\nu_\theta = \frac{1}{\mu} = \frac{\nu}{1+\theta\nu}. \tag{107}$$

Using this, we can easily calculate the **mean signal frequency** ν (i.e. the mean count rate without dead time) from the measured value of the mean count rate ν_θ (i.e. the apparent mean signal frequency) provided that θ is known:

$$\nu = \frac{\nu_\theta}{1-\theta\nu_\theta}. \tag{108}$$

According to the central limit theorem for renewals, N_t is asymptotically $N(t/\mu, t\sigma^2/\mu^3)$ normal. In the case of long enough counting times and constant dead time, therefore, the **expected value** and the **variance of the counts** are:

$$E(N_t) = \frac{\nu}{1+\theta\nu}t = \nu_\theta t, \tag{109}$$

$$D^2(N_t) = \frac{\nu}{(1+\theta\nu)^3}t = (1-\theta\nu_\theta)^2\nu_\theta t. \tag{110}$$

(#70) It follows from above result that:

$$D^2(N_t) = (1-\theta\nu_\theta)^2 E(N_t) \le E(N_t) \tag{111}$$

where the equals sign only holds for $\theta\nu_\theta << 1$. Thus, in the case of measured counts, **the Poisson approximation** (when the variance is taken to be equal to the expected value) tends to **overestimate the actual standard deviation**.

(#71) The product $\theta\nu_\theta$ represents the fraction of the counting time t that is 'covered' by the dead-time intervals associated with the individual measured signals. The **percentage of dead time** displayed by some detecting systems is therefore equal to $100\theta\nu_\theta$.

4.1.3. The 'primeval' shape of the photoelectric peak

The initial stage of the detection is the transferring of the energy of a radiation particle—or a gamma photon—to the substance of the detector. Owing to this process ionized/excited states are produced, the number of which is proportional to the absorbed energy, serving as a basis for the energy determination of the detected particle.

The renewal story can also be recounted with the following exchange of roles: $t \Leftrightarrow E$, $N_t \Leftrightarrow N_E$, $T_i \Leftrightarrow \varepsilon_i$, $\mu \Leftrightarrow \varepsilon$,, $\sigma \Leftrightarrow \sigma_\varepsilon$. Consider, therefore, a gamma photon, the energy (E) of which breaks up to random portions (ε_i) while it is transferred to the detector crystal following photoelectric effect. The renewal process N_E now counts the number of portions—a random

variable—to which the energy of the photon has happened to split. Each energy portion is spent on the production of one charge carrier (or one excited state relevant from the viewpoint of detection). Therefore N_E also means the number of charge carriers produced by the detected photon, which—assuming linear amplification—will be proportional to the pulse height. If further statistical effects can be neglected, then the density function of N_E will be characteristic of the pulse height distribution as well, which has a Gaussian shape according to experience.

The distribution of the ε_i-values is concentrated to the energy range (V, E), where V is the **minimum of energy** (ionization potential), which is **just enough for the production of one charge carrier** in the given detector substance. Further characteristic parameters of the distribution are: $\varepsilon \equiv \langle \varepsilon_i \rangle$, the mean energy spent on the production of one charge carrier/excited state, as well as σ_{rel}, the relative deviation of the ε_i-values.

According to the central limit theorem of renewals, N_E will be asymptotically normal with the following parameters:

$$N\left(\frac{E}{\varepsilon}, \sigma_{\text{rel}}^2 \frac{E}{\varepsilon}\right). \tag{112}$$

The shape of the photoelectric peak is therefore determined by the density function of the above normal distribution (i.e. by a Gaussian curve), the mode/expected value (MAX) and halfwidth (FWHM) of which are:

$$\text{MAX} = \frac{E}{\varepsilon}; \quad \text{FWHM} = 2\sqrt{2\ln 2}\,\sigma_{\text{rel},\varepsilon}\sqrt{\frac{E}{\varepsilon}}\,. \tag{113}$$

Thus the position of the maximum is proportional to the photon energy E, and the low-energy peaks are narrower than the high-energy peaks. We can also see that the relative peak-width is inversely proportional to the square root of the energy:

$$\text{FWHM}_{\text{rel}} \equiv \frac{\text{FWHM}}{\text{MAX}} = 2\sqrt{2\ln 2}\,\sigma_{\text{rel},\varepsilon} \Big/ \sqrt{\frac{E}{\varepsilon}}\,. \tag{114}$$

The above formulae show that this simple stochastic model not only explains the shape of the photoelectric peak, but also predicts that the smaller is the average energy spent on the production of one charge carrier/excited state in a detector, the better is its energy resolution.

4.1.4. Statistical effects smearing the photoelectric peak

We have assumed above that there are no statistical effects at later stages of pulse formation that could change the shape of the photoelectric peak. However, such effects *do* exist. Let us consider the scintillation detectors, for instance, which waste much of the potential of energy resolution hidden in the 'primeval' shape of the photoelectric peak.

In the case of **scintillation detectors**, the production of excited states is practically instantaneous in comparison with the time scale of de-excitation. The excited states are characterized by the same exponential-type lifetime distribution as the radioactive atoms. A large number of excited states start to 'decay' at the same moment ($t = 0$), as a result of which **visible photons** are formed (see FIGURE 20).

The temporal density of photons as a function of time—in other words: the shape of the light pulse—shows the characteristic features of the exponential law of decay. If the

electronics does not mix up the elementary processes too much, then the **shape of the light pulse** will be reflected in the **shape of the voltage pulse** as well, the latter being the object of pulse height analysis. All in all, the pulse height of the detector signal will be determined by the number of photons emitted between $t = 0$ (the moment when the excited states were formed) and a moment a little later: $t = 0 + \Delta t = \Delta t$. (The length of Δt depends on the details of the signal processing/pulse-height analysis, however it is surely shorter than the mean life τ of the excited states.) The number of photons responsible for the pulse height is determined therefore by the $B(N_E, \exp(-\Delta t/\tau))$ binomial distribution for any value of N_E. Since the 'primeval' shape is very narrow, N_E can be replaced by its expected value (MAX). If the binomial expected value MAX × $\exp(-\Delta t/\tau)$ and the variance MAX × $\exp(-\Delta t/\tau)$ × [1 - $\exp(-\Delta t/\tau)$] are not too small, then the binomial distribution can be approximated with normal.

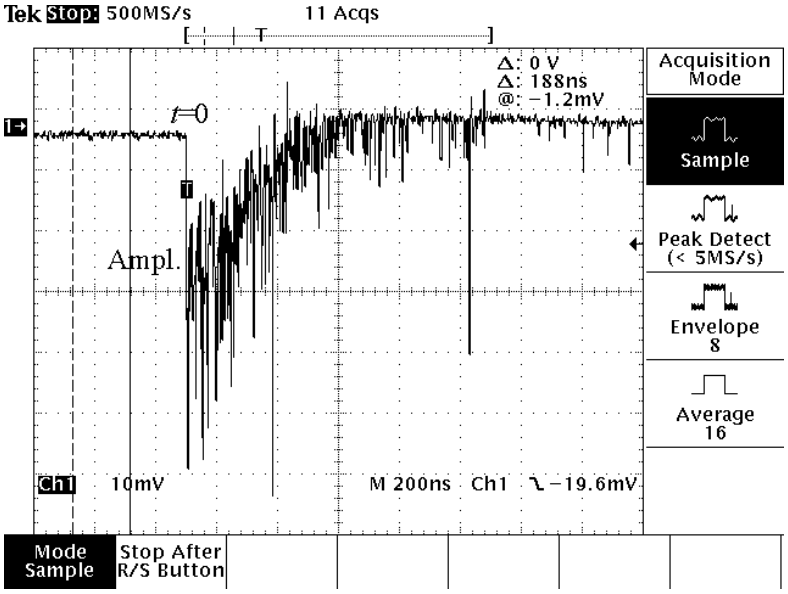

FIGURE 20. **Fine structure of the detector signal** of 0.1 mm NaI(Tl) scintillator at gamma energy of 14.4 keV. The high time-resolution of the digital oscilloscope helps to visualize the elementary processes that make up the signal. The steep rise of the pulse is an indication of the promptness of the formation of the excited states. The individual 'hairs' sticking out from the fuzzy contour represent photoelectrons (or rather the results of cascades started by these in the multiplier) produced by individual visible photons at the photocathode. The multitude of partly merging elementary pulses outlines the **exponential law** of the 'decay' of excited states.

Hence, the number of photons determining the pulse height is normally distributed:

$$N\left(A\frac{E}{\varepsilon}, B\frac{E}{\varepsilon}\right) \tag{115}$$

where A and B are constants. Note that the result is another Gaussian that is considerably broader that the 'primeval' one, but otherwise has the same properties, the most important being the **proportionality between pulse height** (Ampl.) **and energy**:

$$\text{pulse height} \equiv \text{Ampl.} \propto \frac{E}{\varepsilon}. \tag{116}$$

From now on, we will only follow the fate of the photons characterized by the above distribution. The **production of photoelectrons** at the photocathode can be described by

Bernoulli sampling (see at the normal distribution). Bernoulli sampling will broaden the distribution (in the relative sense), however, it will not lead out from the family of normal distributions:

$$N\left(\eta A\frac{E}{\varepsilon},\eta\left[(1-\eta)A+\eta B\right]\frac{E}{\varepsilon}\right)=N\left(a\frac{E}{\varepsilon},b\frac{E}{\varepsilon}\right). \tag{117}$$

The **electron multiplier** will further modify the shape of the peak (see at the **branching processes**), but, fortunately, the proportionality between pulse height and energy will not suffer.

4.2. Markov chains

A Markov chain is a stochastic process defined by the condition:

$$P(a<X_n\le b\big|X_1=x_1,,X_2=x_2,\dots,X_{n-1}=x_{n-1})=P(a<X_n\le b\big|X_{n-1}=x_{n-1}) \tag{118}$$

which means that the current outlook of such a process is determined by the latest information to such an extent that no earlier information can add anything to it.

4.2.1. Branching processes

The story of the branching processes is the following. Consider an entity (0th generation), which produces X_1 identical offspring in one single reproductive cycle (1st generation). The offspring of the nth generation constitute the (n+1)th generation. The entities belonging to each generation proliferate independently of each other and of their predecessors according to the same distribution. We wish to follow X_n, the population of the subsequent generations of entities. Such **entities** can be, e.g. **neutrons multiplied by a chain reaction** or **electrons 'breeding' on the dynodes of a multiplier**.

The X_1 is a discrete random variable that can assume the values k = 0, 1, 2… with different probabilities $P(X_1 = k) = p_k$. Let μ denote the expected value, and σ^2 the variance of the distribution. Let $G(s) \equiv G_1(s)$ be the generating function of the distribution:

$$G(s)\equiv\sum_{k=1}^{\infty}p_k s^k\,. \tag{119}$$

Then the generating functions of the subsequent generations can be calculated from the following recursion formula (Feller 1968):

$$G_{n+1}(s)=G(G_n(s))\,. \tag{120}$$

The expected value and the variance of the branching process are (Feller 1968):

$$E(X_n)=\mu^n\,;\quad D^2(X_n)=(\mu^{2n-2}+\mu^{2n-3}+\ldots+\mu^{n-1})\,\sigma^2\,. \tag{121}$$

Note that the result obtained for the expected value is in accordance with the common-sense expectation.

(#72) FIGURE 21 shows a sequence of distributions resulted from such a calculation. The model can represent, e.g., a **photomultiplier**, the dynodes of which 'double' the incident electrons with a probability of 50%, i.e. half of the electrons have 2 offspring, while the other

half only 1. Thus the mean multiplication factor of the dynodes is $\mu = 3/2$, with a standard deviation of $\sigma = 1/2$.

According to Eq. (121), the **relative deviation of multiplication** is:

$$\sigma_{\mathrm{rel}}(X_n) \equiv \frac{D(X_n)}{E(X_n)} = \frac{\sigma}{\mu}\sqrt{1+\mu^{-1}+\mu^{-2}+\ldots+\mu^{-n+1}}\,. \tag{122}$$

For a **large number of dynodes**, the relative deviation approaches the following limit:

$$\sigma_{\mathrm{rel}}(X_\infty) = \frac{\sigma}{\mu}\frac{1}{\sqrt{1-\mu^{-1}}} = \frac{\sigma}{\sqrt{\mu(\mu-1)}}\,. \tag{123}$$

(#73) Continuing the previous remark: In the case of one stage, the relative deviation is about 33%. By increasing the **number of dynodes,** the relative deviation increases asymptotically to about 58%. The above formula reveals, however, that by increasing the (mean) multiplication factor of the dynode (μ), the relative deviation becomes smaller (for the same σ). If, e.g., the dynodes triple/quadruple the incident electron with equal probability instead of doubling it (i.e. σ remains 1/2, while μ increases from 3/2 to 7/2), then the asymptotic value of the relative deviation will decrease to about 17%.

(#74) We should stress that multiplication goes by electrons. When, e.g., the photocathode emits N electrons as a result of a light pulse, then each electron will be multiplied to a different extent by the dynodes. The total number of electrons collected by the anode is obtained by summation:

$$\Sigma_N = \sum_{i=1}^{N} X_{n,i}\,. \tag{124}$$

Note that we are dealing with a **random sum** here. According to the results discussed at renewals (see Eq. (117)), the random variable N is asymptotically normal with the following parameters:

$$N\left(a\frac{E}{\varepsilon}, b\frac{E}{\varepsilon}\right). \tag{125}$$

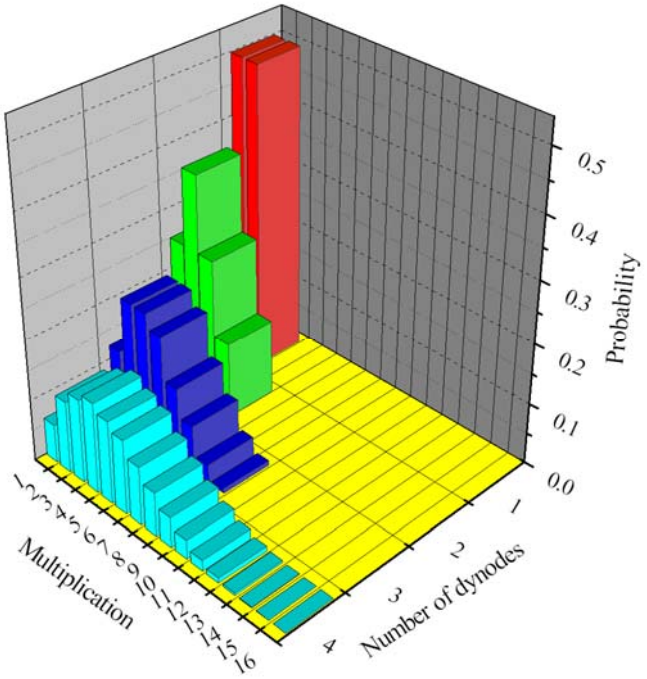

FIGURE 21. Distributions of the multiplication by a '**photomultiplier**' calculated for 1, 2, 3, and 4 dynodes.

According to Eq. (55), the expected value and the variance of Σ_N are as follows:

$$E(\Sigma_N) = \mu^n a \frac{E}{\varepsilon} \propto \frac{E}{\varepsilon}, \tag{126}$$

$$D^2(\Sigma_N) = \left[b\mu^{2n} + a\sigma^2(\mu^{2n-2} + \mu^{2n-3} + \ldots + \mu^{n-1})\right]\frac{E}{\varepsilon} \propto \frac{E}{\varepsilon}. \tag{127}$$

Thus the expected value of the pulse height remains proportional to the energy even after the photomultiplier. (The peak-width also remains proportional to the square root of the energy.)

5. Fitting Nuclear Spectra

In this chapter we will discuss the evaluation of certain parameterized data sequences that we call **nuclear spectra**. For a more general treatment of similar problems we refer to Press *et al.* 1999.

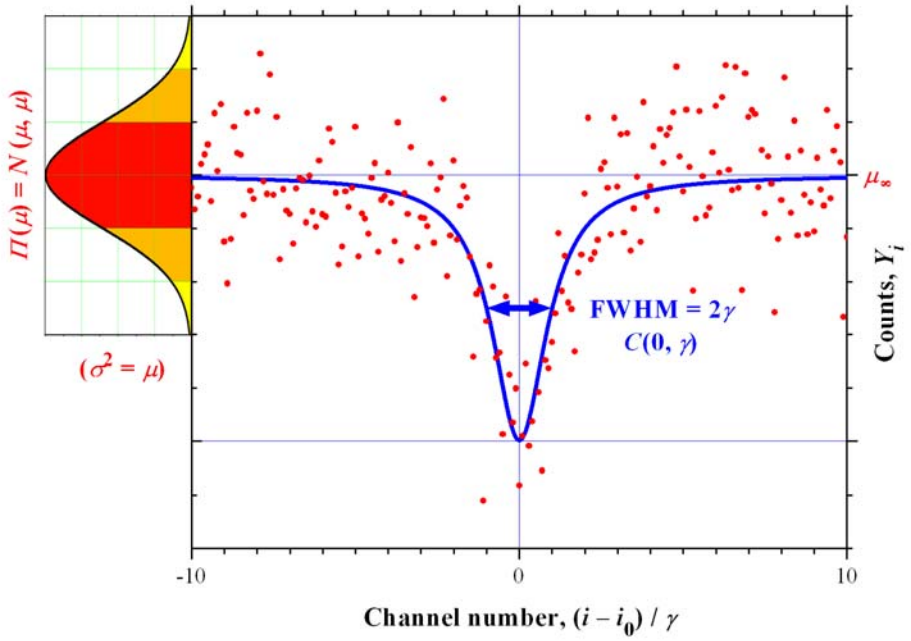

FIGURE 22. An example for a nuclear spectrum. The main graph shows a single-peak Mössbauer spectrum 'measured' at transmission geometry. Such a spectrum can be fitted with a Lorentzian curve (blue line), whose shape is identical with the density function of a Cauchy distribution. Due to standardization, the tick distance on the horizontal axis is half of the FWHM of the Lorentzian (γ). As we have mentioned in remark (#66), FWHM/2 = γ gives the natural line width Γ provided that the absorber is ideally thin. On the other hand, the vertical scattering of the counts (red dots) is characterized by the normal distribution. The colored graph on the left, e.g., shows the normal density function belonging to the baseline (μ_∞). The color code is explained by FIGURE 2. On the vertical axis the distance between the ticks equals to σ.

5.1. Spectra, fitting, model functions

Spectrum. A spectrum is a set of data pairs (spectrum points):

$$(x_i, Y_i) \quad i = 1, 2, \ldots k \tag{128}$$

where x_i is the exact value of some independent variable x (one of a number of set values chosen by the spectroscopist), and the corresponding value of the dependent variable Y_i is considered as a **random variable**. We will assume that the expected value μ_i of Y_i is provided by the appropriate member of a parameterized family of functions

$$\{\mu(x; \boldsymbol{b})\}_{\boldsymbol{b}} \tag{129}$$

called the **model functions**. (The subscript above is to remind that there is an element in the set of functions $\{...\}_{\boldsymbol{b}}$ for each different value of $\boldsymbol{b}$.)

The vector

$$\boldsymbol{b} \equiv (b_1, b_2, \ldots b_n) \tag{130}$$

is called the vector of the **fitting parameters**[18].

Let $\boldsymbol{a}$ denote that particular choice of the parameter vector $\boldsymbol{b}$ at which the fitting actually takes place, i.e. for which:

$$E(Y_i) \equiv \mu_i = \mu(x_i; \boldsymbol{a}) \quad i = 1, 2, \ldots k\,. \tag{131}$$

The function determined by **a**:

$$\mu(x) = \mu(x; \boldsymbol{a}) \tag{132}$$

will be referred to as the **fitting function**. The purpose of fitting is to give the best possible estimate for the parameter vector **a** on the basis of the concrete spectrum. The estimate of **a** will be denoted by $\hat{\boldsymbol{a}}$.

Nuclear spectra. Before sketching out the solution for the fitting problem, we restrict our attention to a special class of spectra called nuclear spectra, in the case of which the Y_is are counts, for which the Poisson approximation and the normal approximation equally hold. In other words, we will concentrate on spectra for which the Y_i-values have $N(\mu_i, \sigma_i^2)$ distributions with $\sigma_i^2 = \mu_i$ as a heritage of the Poisson distribution. (See FIGURE **22**.)

5.2. The maximum likelihood principle

The maximum likelihood principle is based on the following train of thought. Suppose that the shape of the model function is exactly known (in principle), but the exact value of the parameter vector $\boldsymbol{a}$ (providing the fitting function common to all spectra that could be measured under the given conditions) remains hidden from us. We can only state with certainty (because of physical considerations) that such a parameter vector *does* exist. The measured spectrum

$$S_0(\boldsymbol{a}) \equiv \{(x_i, Y_i)_0 \quad i = 1, 2, \ldots k\} \tag{133}$$

can therefore be considered as a concrete member—realized by a concrete measurement—of the manifold of spectra $\{S_\alpha(\mathbf{a})\}_\alpha$ the elements of which are characterized by the same 'exact' parameter vector **a**. The different elements of $\{S_\alpha(\mathbf{a})\}_\alpha$ are not equally likely to occur in an actual measurement. Since the spectrum points have normal distribution and they represent independent random variables, we can assign the following 'probability'—or rather: likelihood—to the different spectra:

$$P(\alpha; \boldsymbol{a}) \Leftrightarrow \prod_{i=1}^{k} \left[\frac{1}{\sqrt{2\pi}\sigma_i} \exp\left(-\frac{1}{2}\left(\frac{Y_i - \mu(x_i; \boldsymbol{a})}{\sigma_i} \right)^2 \right) \Delta Y \right] = \boldsymbol{L}(Y_i; \boldsymbol{a})(\Delta Y)^k \tag{134}$$

where $\boldsymbol{L}(Y_i; \mathbf{a})$ is called the **likelihood function** that can be considered as a k dimensional density function characterizing the 'probability' of the spectrum in question (Orear 1987):

[18] See the example mentioned in remark **(#35)**.

$$\boldsymbol{L}(Y_i;\boldsymbol{a}) = \frac{1}{\prod_{i=1}^{k}\sqrt{2\pi}\,\sigma_i}\exp\left(-\frac{1}{2}\sum_{i=1}^{k}\left(\frac{Y_i-\mu(x_i;\boldsymbol{a})}{\sigma_i}\right)^2\right). \tag{135}$$

Note that the factor ΔY was needed in Eq. (134) so that we could convert the normal density function to probability. Now if we wanted to take a guess at the 'position' of the measured spectrum $S_0(\boldsymbol{a})$ in the manifold $\{S_\alpha(\boldsymbol{a})\}_\alpha$, then the most sensible thing to do would be to look for it around the maximum of the above likelihood. In other words—like other experimenters—we *trust* that the single result of our measurement (the spectrum measured maybe for several days) exemplifies a typical case rather than a rare and extreme one:

$$P(0;\boldsymbol{a}) \approx \max_\alpha P(\alpha;\boldsymbol{a}). \tag{136}$$

Laying this down, we can continue arguing like this. Since we would like to *believe* that the measured spectrum $S_0(\boldsymbol{a})$ is a likely realization associated with the 'real'—alas, unknown—parameter vector $\boldsymbol{a}$, the best thing that we can do is select the estimate $\hat{\boldsymbol{a}}$ so that it (of all the possible values of the parameter vector $\boldsymbol{b}$) assigns the greatest likelihood to the measured spectrum:

$$P(0;\hat{\boldsymbol{a}}) \approx \max_{\boldsymbol{b}} P(0;\boldsymbol{b}). \tag{137}$$

The above maximum condition is an expression of the **maximum likelihood principle**. Note that the maximum likelihood principle—rationality aside—does not provide a guarantee that the 'real' $\boldsymbol{a}$ will be found, since it is based on belief rather than on strict mathematical foundations. However, the results shown below can be interpreted mathematically as well, speaking for the 'strength' of the principle.

5.3. Weighted least squares as a maximum likelihood method

Because of the monotony of the logarithmic function, the above maximum condition can be converted to the following minimum condition:

$$-\log P(0;\hat{\boldsymbol{a}}) \approx \min_{\boldsymbol{b}}(-\log P(0;\boldsymbol{b})). \tag{138}$$

In the concrete situation this condition provides the estimate $\hat{\boldsymbol{a}}$ as the solution of the

following minimization:

$$\chi^2 \equiv \sum_{i=1}^{k}\left(\frac{Y_i-\mu(x_i;\hat{\boldsymbol{a}})}{\sigma_i}\right)^2 \Rightarrow \text{minimum} \tag{139}$$

Note that the above sum of squares is proportional to the weighted average of the squared deviation from the expected value as defined by Eq. (36).

There are commercial programs for solving the above type of minimization. They provide not only the estimates of the unknown parameters but also their variances and the correlation coefficients between them. Instead of going into details of such evaluations, we only make a couple of remarks concerning the minimum of χ^2 that is characteristic of the **goodness of fit**.

(#75) The notation χ^2 is not accidental. We have seen at the χ^2 distribution that when the model function is linear in the fitting parameters, then the minimized sum of squares has a $\chi^2(k-n)$ distribution, i.e. the goodness of fit can be judged accordingly.

(#76) If the model function is not linear in all of the parameters, then, perhaps, it can be linearized by a transformation or using a power series. After linearization, the sum of squares will certainly have a χ^2 distribution with *some* degree of freedom. In order to find the minimum, however, the model function need not be linearized (Press *et al.* 1999). The possibility of linearization is only necessary for the declaration of the 'competence' of the χ^2 distribution as a goodness-of-fit measure, as well as for finding the value of n pointing to the concrete χ^2 distribution 'in charge'. Considering that the value of k can be as large as several thousand in the case of nuclear spectra, the value of n (which is much less than k) does not matter too much, because the sum of squares will approximately have a $\chi^2(k-n) \approx \chi^2(k) \approx N(k, 2k)$ distribution anyway.

(#77) Some types of linearization require the (non-linear) transformation of the measured spectrum. As an example, we mention the logarithmic conversion of decay data for the determination of the decay constant. Logarithmic conversion spoils the initially normal distribution of the spectrum points, thus breaking the connection with the χ^2 distribution as a goodness-of-fit measure.

5.4. Weighted least squares method in nuclear spectroscopy

We have not yet utilized that $\sigma_i^2 = \mu_i$ for nuclear spectra. Before including this relationship in the least squares method, we will consider a very simple nuclear spectrum consisting of the counts measured with a long-lived radionuclide for equal periods of time. The model function is obviously very simple in this case. It only contains one single parameter representing the common expected value of the measured counts. The minimization problem therefore can be expressed like this:

$$\chi^2 \equiv \sum_{i=1}^{k} \left(\frac{Y_i - \hat{a}}{\sigma_i} \right)^2 \Rightarrow \text{minimum} \tag{140}$$

where $\hat{a}$ is the estimate of the constant $\mu = E(Y_i)$. In this particular case there are three—equally sensible—choices for giving the value of σ_i^2. The question is, whether or not they lead to the same estimate.

Case 1: $\sigma_i^2 = \sigma^2 =$ constant. We have only used that the points have the same variance:

$$\chi^2 \equiv \sum_{i=1}^{k} \left(\frac{Y_i - \hat{a}}{\sigma} \right)^2 \Rightarrow \text{minimum}\,. \tag{141}$$

Case 2: $\sigma_i^2 = \sigma^2 = \mu \approx \hat{a}$. Here, we have utilized that we are dealing with a nuclear spectrum, and also that $\hat{a}$ is the estimate of μ:

$$\chi^2 \equiv \sum_{i=1}^{k} \frac{(Y_i - \hat{a})^2}{\hat{a}} \Rightarrow \text{minimum}\,. \tag{142}$$

Case 3: $\sigma_i^2 = \sigma^2 = \mu \approx Y_i$. Here, we have utilized that we are dealing with a nuclear spectrum, and also that, in the case of Poisson distribution, a single measured value is a fair

estimate of the expected value[19], provided that the measured (and thus the expected) value is large enough:

$$\chi^2 \equiv \sum_{i=1}^{k} \frac{(Y_i - \hat{a})^2}{Y_i} \Rightarrow \text{minimum}. \tag{143}$$

Differentiating the above expressions with respect to $\hat{a}$, we find that each minimum problem has a different solution. The result will always be some type of an average of the Y_i-values, however, in case 1 it is the **arithmetical mean** (i.e. the sample mean, see Eq. (29)), in case 2 the **harmonic mean** (see Eq. (19)), while in case 3 the **root mean square** (see Eq. (17)), what we get.

And now let us return to the problem in general. However tempting **case 1** seems to be (note that it proved to be related to the sample mean that is both the unbiased estimate of the expected value and, as we have just seen, its maximum likelihood estimate), it cannot serve as a general model, because the model function is not usually constant and therefore Eq. (141) does not represent a maximum likelihood condition in general.

Although **case 2** conveys the nuclear character of the spectra quite accurately, it has two major drawbacks. First, the derivatives are more complicated and, second, the statistical interpretation is more difficult. Consider, e.g. that all the advantages of a linear model function disappear as soon as that same model function appears in the denominator as well (see Eq. (142)).

In **case 3** the nuclear character is only expressed approximately, but the approximates (Y_i) can be considered as constants from the viewpoint of differentiation and therefore the following minimum condition/**merit function** is accepted in the practice of spectrum evaluation:

$$\chi^2 \equiv \sum_{i=1}^{k} \frac{(Y_i - \mu(x_i; \hat{\boldsymbol{a}}))^2}{Y_i} \Rightarrow \text{minimum} \tag{144}$$

As regards the relationship between χ^2-values and **confidence intervals** we refer to the excellent literature available (Press *et al*. 1999) and we content ourselves with a few remarks only.

(#78) For rather obvious reasons, the following relationships hold between the above-mentioned means:

$$\textbf{harmonic mean} \leq \textbf{sample mean} \leq \textbf{root mean square} \tag{145}$$

It is reassuring, however, that in the case of the Poisson distribution the actual difference between the above 'averages' is relatively small. It is easy to check (e.g. with random numbers of $\Pi(\mu)$ distribution) that the difference—varying μ between 10^2 - 10^6 and 'averaging' ten data at a time—does not usually exceed 1. On the other hand, for $\mu = 10^2$, 10^4, and 10^6 the standard deviations are 10, 100, and 1000, respectively, showing that the three methods of 'averaging' are practically equivalent, because the standard deviations are much larger than the difference between them.

(#79) In spite of the fact that the χ^2-value is the result of minimization, the same 'protocol' should be followed as usual, i.e. when the confidence/acceptability of an estimate is judged by using any other distribution as a measure. Therefore, the fit giving the smallest χ^2-value is not necessarily the best. The best fit is the one whose χ^2-value is nearest to the expected value (determined by the degrees of freedom) of the corresponding χ^2 distribution. It is true,

[19] See remark **(#24)**.

however, that in 99 cases out of 100 we are unsatisfied with the fit because we find the χ^2-**value too large**.

One possible reason for this can be that the **model function used is unsatisfactory**. It is a serious error, e.g. when the dimension of the vector $\boldsymbol{b}$ is taken too small (this happens, e.g. when we try to fit a Mössbauer spectrum with fewer peaks than we should). Another source of such error can be that the shape of the elementary model function is imperfect (e.g. the peaks of a Mössbauer spectrum cannot always be described perfectly by Lorentzians).

Spectrum points that are **way out of range** (mainly: 'dropped' points) can also lead astray the fitting process.

(#80) If the **relative χ^2 is too small**, it can be an indication that the fitting program tends to overestimate the degrees of freedom. This can happen, e.g., when **the model function is not linear in each of the parameters**, and the degree of freedom is calculated (according to the general practice) simply by subtracting the number of fitted parameters from the number of spectrum points.

Too small (i.e. too good) a χ^2 can be obtained, when the **dead time** is so large that the Poisson approximation built into the **merit function** of Eq. (144) overestimates the variance in the denominator. In such a case, we get a more realistic result, if—on the basis of Eq. (111)—we divide the calculated χ^2-value by $(1-\theta\nu_\theta)^2$, where $\theta\nu_\theta$ is the percentage of dead time divided by 100.

6. Summarizing Tables

TABLE 1. **Addition theorems.** The addition theorems summarized below are only valid if the random variables X_1 and X_2 are independent of each other.

Distribution of X_1	Distribution of X_2	Distribution of $X_1 + X_2$
$B(n_1, p)$	$B(n_2, p)$	$B(n_1+n_2, p)$
$\Pi(\mu_1)$	$\Pi(\mu_2)$	$\Pi(\mu_1+\mu_2)$
$\gamma(r_1, \nu)$	$\gamma(r_2, \nu)$	$\gamma(r_1+r_2, \nu)$
$N(\mu_1, \sigma_1^2)$	$N(\mu_2, \sigma_2^2)$	$N(\mu_1+\mu_2, \sigma_1^2+\sigma_2^2)$
$\chi^2(k_1)$	$\chi^2(k_2)$	$\chi^2(k_1+k_2)$
$C(m_1, \gamma_1)$	$C(m_2, \gamma_2)$	$C(m_1+m_2, \gamma_1+\gamma_2)$

TABLE 2. **Limiting distributions.** The conditions given below are practical substitutes for the exact limits (such as $p \to 0$ and $n \to \infty$ in the case of the first row). Under the practical conditions given in the middle column, the limiting distribution given in the third column gives a fair approximation of the distribution given in the first column.

Distribution of X	Conditions	Limiting distribution of X
$B(n, p)$	$p \le 0.1$ $n \ge 20$	$\Pi(np)$
$B(n, p)$	$npq \ge 6$ $q \equiv (1-p)$	$N(np, npq)$
$\Pi(\mu)$	$\mu \ge 20$	$N(\mu, \mu)$
$\gamma(r, \nu)$	$k \ge 30$	$N(r/\nu, r/\nu^2)$
$\chi^2(k)$	$k \ge 50$	$N(k,2k)$

TABLE 3. **Normal distribution**
Probabilities expressed in percent for deviations from the mean higher than $\mu \pm d\sigma$ for the random variable X having $N(\mu, \sigma^2)$ normal distribution.

d	$100\times P(\lvert X-\mu\rvert \geq d\sigma)$	d	$100\times P(\lvert X-\mu\rvert \geq d\sigma)$
0.67449	50	2.1	3.57
0.7	48.39	2.2	2.78
0.8	42.37	2.3	2.14
0.9	36.81	2.4	1.64
1.0	31.37	2.5	1.24
1.1	27.13	2.6	0.932
1.2	23.01	2.7	0.693
1.3	19.36	2.8	0.511
1.4	16.15	2.9	0.373
1.5	13.36	3.0	0.27
1.6	10.96	3.5	0.0465
1.7	8.91	4	0.00634
1.8	7.19	5	0.0000573
1.9	5.74	6	0.0000002
2.0	4.55	7	0.00000000026

TABLE 4. **χ^2 distribution**
The p-quantiles of $\chi^2(k)$ distributions for different degrees of freedom k

k\p	**0.01**	**0.05**	**0.10**	**0.25**	**0.50**	**0.75**	**0.90**	**0.95**	**0.99**
1	0.0002	0.0039	0.0148	0.102	0.455	1.32	2.71	3.84	6.33
5	0.554	1.15	1.61	2.67	4.35	6.63	9.24	11.1	15.1
10	2.56	3.94	4.87	6.74	9.34	12.5	16.0	18.3	23.2
15	5.23	7.26	8.55	11.0	14.3	18.2	22.3	25.0	30.6
20	8.26	10.9	12.4	15.5	19.3	23.8	28.4	31.4	37.6
25	11.5	14.6	16.5	19.9	24.3	29.3	34.4	37.7	44.3
30	15.0	18.5	20.6	24.5	29.3	34.8	40.3	43.8	50.9

Acknowledgements

The author gratefully acknowledges Dr. Károly Süvegh's help for kindly doing the positron lifetime measurements for FIGURE 13. He also thanks Dr. György Vankó, his former colleague, for giving permission to use his diagram in FIGURE 20 as well as Prof. György Michaletzky of the Department of Probability Theory and Statistics, for helpful advice.

References

BELGYA, T., FAZEKAS, B., MOLNÁR, G., GATENBY, R. A., JOHNSON, E. L., BAUM, E. M., WANG, D., DIPRETE, D. P., YATES, S. W.,1993, *Eighth Intern. Symp. on Capture Gamma-Ray Spectroscopy.*,
Fribourg, Switzerland (20-24 September 1993), pp. 878-887, FIGURE 4.

BEVINGTON, P. R., 1969, *Data Reduction and Error Analysis for the Physical Sciences* (New York: McGraw-Hill Book Co.).

COHEN, E. R., GIACOMO, P., 1987, *Physica A*, **146**, pp. 1-68.

FELLER, W., 1968, *An Introduction to Probability Theory and its Applications*, Vol. I, 3rd edition,
(New York: John Wiley and Sons, Inc.).

GEORGII, H-O., 2008, *Stochastics: introduction to probability and statistics* (Berlin: de Gruyter).

GOODMAN, R., 1988, *Introduction to Stochastic Models* (Menlo Park, Amsterdam, Sydney, Tokyo, Madrid: The Benjamin/Cummings Publishing Company, Inc.).

HOGG, R. V., TANIS, E. A., 1988, *Probability and Statistical Inference*, 3rd edition, (New York: Macmillan Publishing Company).

JÁNOSSY, L., 1965, *Theory and Practice of the Evaluation of Measurements* (Oxford: Clarendon Press).

KARLIN, S., TAYLOR, H. M., 1975, *A First Course in Stochastic Processes* (New York, San Francisco, London: Academic Press).

KARLIN, S., TAYLOR, H. M., 1984, *An Introduction to Stochastic Modeling* (Orlando: Academic Press).

KNOLL, G. F., 1979, *Radiation Detection and Measurement* (New York: John Wiley and Sons, Inc.), Chapter 4.

KORN, G. A., KORN, T. M., 1968, *Mathematical Handbook for Scientists and Engineers*, 2nd edition
(New York: McGraw-Hill Book Company), Chapters 18-19.

LEO, W. R., 1987, *Techniques for Nuclear and Particle Physics Experiments* (Berlin: Springer-Verlag), Chapter 4.

LUX, J., KOBLINGER, L., 1991, *Monte Carlo Particle Transportation Methods: Neutron and Photon Calculations* (Boston: CRC Press).

LYONS, L., 1986, *Statistics for Nuclear and Particle Physicists* (Cambridge: University Press).

OREAR, J., 1987, *Notes on Statistics for Physicists*. In *Procedures Manual for the Evaluated Nuclear Structure Data File*, edited by M. R. Bhat, (National Nuclear Data Center Brookhaven National Laboratory).

PRESS, W. H., TEUKOLSKY, S. A., VETTERLING, W. T., FLANNERY, B. P., 1999, *Numerical Recipes in C—The Art of Scientific Computing*, 2nd edition (Cambridge: University Press), Chapters 14-15.

RÉNYI, A., 1970, *Foundations of Probability* (San Francisco, Cambridge, London: Holden-Day, Inc.).

RÉNYI, A., 1979, *Wahrscheinlichkeitsrechnung*. *Mit einen Anhang über Informationstheorie*, 6. Aufl.
(Berlin: VEB Deutscher Verlag der Wissenschaften).

SCHMIDT, K.-H., SAHM, C. C., PIELENZ, K. CLERC, H. G., 1984, *Z. Phys. A*, **316**, 19.

Index

#

A

B

C

D

E

R

S

T

U

V

W