Appendix
Statistics and Probability

Abstract Statistics and probability distributions are an important aspect of radiation interaction processes, owing to the intrinsic stochastic nature of interactions. This Appendix is intended to collect in one place a number of related definitions, concepts and formulae on statistics and probability distributions. The style is very compact but the four sections on probability and probability distribution functions can be used to recall the formalism, basic statistics and the mathematical form of common probability distributions.

A.1 Sample Space and Probability

Consider a certain phenomenon that might or might not happen. That is, the occurrence of such phenomenon could be one of three situations: *certain*, meaning the process does happen with certainty; it could be *impossible*, that is it does not happen at all; or *random*, meaning it might or might not happen. Examples of such phenomena can be recognised from throwing a dice: if we are interested in the outcome corresponding to the draw of any number between 1 and 6, then the phenomenon is certain. It is certain that in throwing a standard dice the outcome will be some number between 1 and 6. On the other hand, it is impossible to draw 0 from a standard dice, so this would be an example of an impossible outcome. Finally, if the outcome of interest is defined as obtaining the number 3 from one single trial, then such phenomenon is random. Statistics deals with random processes giving them a degree of predictability. The *Sample Space* is defined as the set of all possible outcomes of an experiment, a trial, or a phenomenon. The *Event* is defined as a subset of the sample space, namely it is a set of outcomes. Let *Elementary Event* be defined as one single outcome in the sample space. A couple of examples should help place in focus such definitions. Two coins are tossed and such trial could lead to four possible outcomes: two tails, two heads, or 1 tail and 1 head with the possible permutation of the order. These four outcomes define the sample space; each outcome is an elementary event, while the outcomes with at least one head (3 elementary events) define an event.

L. Cerrito, *Radiation and Detectors*, Graduate Texts in Physics,
DOI 10.1007/978-3-319-53181-6

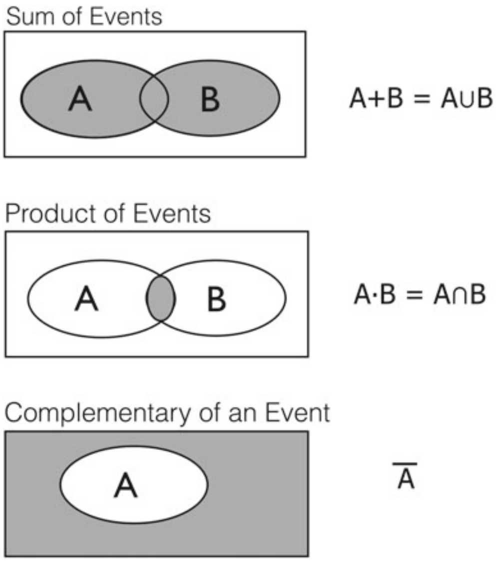

Fig. A.1 Algebra operations of sample spaces A and B into C, showing the sum, the product and the complementary. In shaded colour is indicated the set of elementary events belonging to the resulting event space, C

A second example is given by rolling one dice; the possible outcomes are integer numbers between 1 and 6, therefore such set defines the sample space. If the outcome of interest is the drawing of any number greater than 3, this defines an event, composed of three elementary events (the numbers 4, 5 and 6).

It is possible to define an Algebra of events, with common operations defined as follows. The space C, *Sum* of two sample spaces, A and B, is defined as the set of elementary events that belong to A or B. The space C, *Product* of two sample spaces, A and B, is defined as the set of elementary events that belong to A and B. The *Complementary* of a sample space A (indicated with $\bar{A}$) is the set of elementary events that do not belong to A. In mathematical formalism (see also Fig. A.1):

$$C = A + B = A \cup B, \tag{A.1}$$

$$C = A \cdot B = A \cap B. \tag{A.2}$$

The definition of Probability is not unique and there are several ways to approach it. There are at least three classes of definitions which are mostly relevant. In the so-called *classical* interpretation, considering a sample space with N elementary events and an event A consisting of a number of elementary events, n, then the probability of A is defined as the ratio

$$P(A) = \frac{n}{N}. \tag{A.3}$$

For example, the probability for obtaining the number 3 from a fair dice would be, $n = 1, N = 6, P(A) = 1/6$. There are clear limitations to such definition. It assumes a finite number of outcomes and it considers only elementary events as being equally *likely*, besides the problem that the meaning of such likelihood is not itself defined.

A better definition is given by the *frequentist* interpretation of probability. Considering an event A part of a sample space, the frequency of A is defined as

$$F(A) = \frac{m}{M}, \tag{A.4}$$

where m is the number of cases in which A is obtained in a repetition of the outcomes and M is the number of repetitions, or trials. The probability of A is then defined as the limit of the frequency of A for infinite repetitions, namely

$$P(A) = \lim_{M \to \infty} F(A). \tag{A.5}$$

Compared to the classical interpretation of (A.3), the frequentist definition of (A.5) does not require equally likely elementary events, nor the concept of likelihood itself, but it does need an infinite number of repetitions in order to define the probability.

The third class of probability definitions is the *axiomatic basis* (Kolmogorov 1933 [1]), which can be summarised in the following statements: given a sample space S and events A with $A \in S$ (the symbol $\in$ means that A belongs to S):

- For any A, there is one real number $\mathrm{P}(A)$ associated with A such that $P(A) \geq 0$,
- $\mathrm{P}(S) = 1$ and $\mathrm{P}(\oslash) = 0$, where $\oslash$ indicates the impossible event
- If $A \cap B = \oslash$ (the events are mutually exclusive) then $\mathrm{P}(A + B) = \mathrm{P}(A) + P(B)$

These definitions allow deriving a number of important properties routinely used in statistics, whose proof goes beyond the scope of this Appendix, for instance

- if C is the certain event, $\mathrm{P}(C) = 1$
- $\mathrm{P}(\bar{A}) = 1 - \mathrm{P}(A)$
- $\mathrm{P}(A \cup \bar{A}) = 1$.

Moreover, if A and B are compatible and overlapping events then the probability of their sum is

$$P(A + B) = P(A) + P(B) - P(A \cdot B), \tag{A.6}$$

where $A \cdot B$ is the overlapping part of the event.

Conditional probability is also particularly important, calculated as

$$P(A|B) = \frac{P(A \cdot B)}{P(B)}, \tag{A.7}$$

where $P(A|B)$ indicates the probability of event A *after* event B has occurred or else within the set of elementary events belonging to event B. The expression in (A.7) is general but if the events A and B are independent then by definition of independent events the conditional probability of A is simply the probability of A, therefore

$$P(A|B) = P(A) \quad \text{(independent events)}, \tag{A.8}$$

since event A cannot bear any knowledge of B. By comparing (A.7) with (A.8) if follows a useful expression for calculating the probability of overlapping independent events

$$P(A \cdot B) = P(A) \cdot P(B) \quad \text{(independent events)}. \tag{A.9}$$

A.2 Probability Distribution Function

Consider a sample space consisting of a set of N events x_i $(i = 1, \ldots, N)$, each with probabilities P(x_i). The collective probabilities P(x_i) are the *probability distribution function* (PDF) of the sample space under consideration. For example, given a sample space of 100 individuals, each being born in one of 4 quarters, January-March (Q_1), April-June (Q_2), July-September (Q_3), October-December (Q_4), the probabilities associated with the birth in each quarter (the events), indicated by P(Q_i) is a probability distribution function. An histogram with $Q_{1,\ldots,4}$ on the horizontal axis and P(Q_i) on the vertical axis would be a graphical representation of this probability distribution function.

Two quantities are extremely useful in the description of a PDF defined as above. Consider once more the sample space X consisting of a set of N events with values x_i, $(i = 1, .., N)$ with probabilities P(x_i). The first quantity of interest is the *expectation value*, indicated with $E[X]$

$$E[X] = \sum_{i=1}^{N} x_i P(x_i), \tag{A.10}$$

that is the sum of the products of the event value x_i multiplied by its probability. Notice how (A.10) resembles the definition of average value of a repeatable sequence

$$\bar{x} = \frac{1}{N} \sum_{i=1}^{N} x_i, \tag{A.11}$$

and it is in fact a conceptual extension of the same idea (or rather it justifies the definition of average), representing the "center of gravity" of the distribution.

The second quantity useful to describe a PDF is the *variance*, defined as

$$var[X] = E\left[(X - E(X))^2\right], \tag{A.12}$$

which addresses the spread of the distribution. This resembles the definition of root-mean-square (RMS) of a sequence of N numbers,

$$RMS^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2 \tag{A.13}$$

and this is conceptually the same idea (or, rather, the variance underpins the definition of root-mean-square). It is easy to show that

$$var[X] = E[X^2] - E^2[X], \tag{A.14}$$

which is sometimes a simpler expression to compute than what given in (A.13).

A.3 Binomial and Poisson Distributions

The PDF described in Sect. A.2 are discrete probability distribution functions, meaning that the N events take a discrete set of values x_i. An example of such distribution was discussed with the probability of the quarter of birth in a given population. Some PDF are however so common that have been studied in detail. One such PDF is the *Binomial* probability distribution function,

$$P_{N,p}(n) = \frac{N!}{n!(N-n)!}p^n(1-p)^{N-n} \quad \text{(Binomial distribution)}, \tag{A.15}$$

where N and n are integers, N is the total number of events, n is the value each event takes and p is a parameter of the distribution (a real number). Statistical processes described by a binomial distribution are extremely common, for instance (A.15) describes the probability of obtaining n tails from tossing one coin N times (or equivalently N coins once) and p is the single-trial probability. It can be shown that from the definitions of expectation value and variance applied to the distribution of (A.15), the binomial distribution has the following properties,

$$E[n] = \sum_{n=0}^{N} n P_{N,p}(n) = Np, \tag{A.16}$$

$$var[n] = Np(1-p). \tag{A.17}$$

A discrete PDF of exceptional importance is also the *Poisson* distribution; given a total of N events, the probability of each n event is given by

$$P_\lambda(n) = \frac{\lambda e^{-\lambda}}{n!} \quad \text{(Poisson distribution)}, \tag{A.18}$$

where λ is a real number and a parameter of the distribution. The expectation value and the variance of the Poisson distribution can be derived from the definitions given in (A.10) and (A.12), yielding

$$E[n] = \lambda, \tag{A.19}$$

$$var[n] = \lambda. \tag{A.20}$$

The Poisson distribution describes for instance the number of radioactive decays from a large number of atoms, given a fixed decay probability $p = \lambda/N$. Moreover, the mathematical expression of (A.18) can be derived as the limit of a binomial distribution for large N ($N \rightarrow \infty$), small probability p ($p \rightarrow 0$) but with their product being a finite number, $Np = \lambda$.

A.4 Continuous Probability Distribution Functions

The probability distributions functions listed in Sect. A.3 are built on an event space consisting of integers (for instance the number of times a tossed coin yields a head, the radioactive decay counts, etc.). The event space can be also a continuum of real numbers, x, and the probability of the events X is defined in such case as

$$P(x < X < x + \mathrm{d}x) = f(x)\mathrm{d}x, \tag{A.21}$$

where $f(x)$ is called the *Probability Density Function*. Clearly, the probability density function must be integrable,

$$\int_{-\infty}^{+\infty} f(x)\mathrm{d}x = 1, \tag{A.22}$$

in order to satisfy the condition that the total probability across the full event space, $P(S)$, equals 1.

A simple example of a continuous probability density function is the *Uniform* distribution, which is a constant in the range of real numbers between boundaries a and b

$$f(x) = \text{constant} = K \quad \text{(Uniform distribution)}, \tag{A.23}$$

where K is the constant value and a parameter of the distribution. The condition of (A.22) yields the following relationships

$$\int_{-\infty}^{+\infty} f(x)\mathrm{d}x = 1 \Rightarrow K(b-a) = 1 \tag{A.24}$$

$$K = \frac{1}{b-a}, \tag{A.25}$$

and therefore a uniform distribution between boundaries a and b has form

$$f(x) = \text{constant} = \frac{1}{b-a}. \tag{A.26}$$

The expectation value and the variance can be derived from the definitions given in (A.10) and (A.12), yielding

$$E[x] = \int_a^b x f(x) \mathrm{d}x = \frac{1}{2}(b + a)\,, \tag{A.27}$$

$$var[x] = \int_a^b (x - E[x])^2 f(x) \mathrm{d}x = \frac{1}{12}(b - a)^2\,. \tag{A.28}$$

However, arguably the most important probability density function is the *Gaussian* distribution

$$G(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad \text{(Gaussian distribution)}, \tag{A.29}$$

where σ and μ are parameters of the function (real numbers) and the normalisation condition is

$$\int_{-\infty}^{+\infty} G(x) \mathrm{d}x = 1. \tag{A.30}$$

The Gaussian function has the characteristic bell shape and some of its features are directly derived from the expectation value and the variance:

$$E[x] = \int_{-\infty}^{+\infty} G(x) dx = \mu \tag{A.31}$$

$$var[x] = \int_{-\infty}^{+\infty} (\mu - E[x])^2 G(x) dx = \sigma^2, \tag{A.32}$$

so the curve is centered around μ and its width is parametrised by σ. The Full Width at Half Max (FWHM), sometimes also a useful quantity, is positioned at values of x

$$x = \mu \pm \sigma\sqrt{2\ln 2} \simeq \mu \pm 1.17\,\sigma. \tag{A.33}$$

The integrals of the curve of (A.29) between two arbitrary edges give the probability intervals for the Gaussian distribution, and can be obtained in statistic tables. Most useful numbers are the integrals between one σ around the mean (called one standard deviation), which corresponds to an area of 68% of the total, or 0.68 if the Gaussian function is normalised to unit area. The area between two or three standard deviations around the mean corresponds to 95.5 and 99.7%, respectively. Finally, it can be shown that a Poisson distribution tends to a Gaussian shape in the large λ limit, and a Binomial distribution also tends to a Gaussian shape in the large N limit. This is part of a more general behaviour known as the *Central Limit* theorem.

One final PDF that is worth mentioning is the Chi-squared distribution

$$f(\chi^2) = \frac{1}{2^{n/2}\Gamma\left(\frac{n}{2}\right)} \left(\chi^2\right)^{(n/2-1)} e^{-\chi^2/2} \quad \text{(Chi} - \text{squared distribution)}, \tag{A.34}$$

where χ^2 is the continuous event variable, n is a parameter of the distribution (also known as the degrees of freedom) and Γ is a Gamma function. This distribution is approached by the squared sum of variables with Gaussian distributions and has expectation value and variance:

$$E[\chi^2] = n \tag{A.35}$$

$$var[\chi^2] = 2n \tag{A.36}$$

Furthermore, for large values of n the Chi-squared tends to a Gaussian shape.

A.5 Multivariate and Estimators

The probability distribution functions described so far are relative to one quantity, x, in the event space to which it is associated a probability or a probability density. Consider now two or more quantities, x and y each with its own PDF, $P_X(x)$ and $P_Y(y)$, where $P_X(x)$ and $P_Y(y)$ can be different. The event defined by the sum $z = x + y$ will be characterised by a PDF, indicated by $P_Z(z)$ which is not known apriori. However, given a linear combination of N variables x_i independent and characterised by the same, unknown probability distribution function, the sum z

$$z = a_1 x_1 + a_2 x_2 + a_3 x_3 + \ldots, \tag{A.37}$$

where a_i are real numbers, has the property that for N growing to infinity the probability distribution function of z tends to a Gaussian. This behaviour makes the Gaussian distribution particularly important in experimental physics. It implies that the sum of many independent measurements with the same probability distribution (not necessarily Gaussian) has approximately a Gaussian distribution. If N is not large then z does not have a Gaussian PDF, however the expectation value of the sum of two random variables, x and y, is

$$E[z] = E[x] + E[y], \tag{A.38}$$

and the variance, provided x and y are independent, is

$$var[z] = var[x] + var[y]. \tag{A.39}$$

The (A.38) and (A.39) underly the expressions used in the treatment of experimental data when dealing with repeated measurements of the same quantity x, each measurement being interpreted as a random variable. The expectation value of the mean (A.11) is the same as the expectation value of each measurement of x, while the variance of the mean is related to the variance of each measurement as

$$E[\bar{x}] = \frac{1}{N} E\left[\sum x_i\right] = E[x] \tag{A.40}$$

$$var[\bar{x}] = var\left[\frac{1}{N}\sum x_i\right] = \frac{1}{N} var[x] \tag{A.41}$$

Moreover, in the estimate of the properties of a distribution from a data sampling, the mean $\bar{x}$ is an estimator for the expectation value of the probability density function of x

$$\bar{x} \rightarrow E[x] \tag{A.42}$$

and the RMS is an estimator for the variance

$$RMS_B^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2 \rightarrow var[x], \tag{A.43}$$

where (A.43) differs from (A.13) only for the $N - 1$ in the denominator (Bessel correction).

Reference

1. A. Kolmogorov, *Grundbegriffe der Wahrscheinlichkeitsrechnung* (Springer, Berlin, 1933). English translation by N. Morrison Foundations of the Theory of Probability (Chelsea, New York) in 1950

Index

L. Cerrito, *Radiation and Detectors*, Graduate Texts in Physics,
DOI 10.1007/978-3-319-53181-6