

Random Variables

The idea of a random variable or random event involves no assumption about the intrinsic nature of the phenomenon under investigation. Indeed, it may be a perfectly deterministic phenomenon, and yet a description of the measured quantities in terms of random variables can be extremely productive. If a mathematical theory is judged by its results, there can be no doubt that this is a very useful approach. In this book, we shall discuss this theory in detail, concentrating mainly on results that are relevant to applications in physics.

To fix ideas, let us illustrate the notion of a random variable by an example from everyday life. Time is a physical quantity whose deterministic nature is not difficult to accept. Suppose, however, that you are in the countryside without a watch and that the time is 16h37. It would be very difficult to estimate the time so precisely. On the other hand, you might be able to say that there are 90 chances out of 100 that the time is between 16h00 and 17h00; and if it is winter, that there are 100 chances out of 100 that the time is between 8h00 and 20h00, as there is rarely daylight during the night time. In other words, although the exact time can be considered as deterministic, the quantity to which one has access experimentally will only be an estimate. Now any estimate can be tainted with uncertainty, or noise as the engineers call it, which one may seek to characterize using the theory of probabilities.

We see therefore that, in the phenomenological context in which we find ourselves, the aim is not to investigate the intrinsic nature of the objects concerned, but rather to build up techniques using only the information available to us. We thus adopt this standpoint with regard to the relevant events, relaxing our hypotheses about the specific character of the events themselves. We do not pay attention to the essence of the object, but concentrate on the measurements and predictions we may make, a practice that has led to progress in a great many areas of physics.

2.1 Random Events and Probability

We begin with the straightforward observation of events which we describe as random in order to express the fact that we do not know what will be observed. Consider first the simple case where the set Ω of possible random events is finite, i.e., it contains a finite number of elements. This is the case, for example, for the set of possible outcomes on a lottery wheel. Suppose that this set Ω contains N possible events λ_i , where the index i takes values from 1 to N , so that $\Omega = \{\lambda_1, \dots, \lambda_N\}$. This set Ω can be made up of quite arbitrary elements, with no particular mathematical structure. (The population described by the set is then said to be amorphous.) We may assign a number p_i between 0 and 1 to each event λ_i . This set of N positive numbers will be called a probability law on Ω if $p_1 + \dots + p_N = 1$, or written more succinctly,

$$\sum_{i=1}^N p_i = 1 .$$

We then say that p_i is the probability of λ_i and we shall write $p_i = P(\lambda_i)$.

In the case where the set is infinite but countable, in the sense that the elements can be numbered by the positive integers, these ideas are easily generalized. We write $\Omega = \{\lambda_1, \lambda_2, \dots, \lambda_n, \dots\}$ and

$$p_1 + p_2 + \dots + p_n + \dots = 1 \quad \text{or} \quad \sum_{i=1}^{\infty} p_i = 1 .$$

With this definition of probability, it is sometimes possible to identify the probability with the frequency of occurrence of the relevant event. Consider the trivial example of tossing a coin. In this case, the two possible random events are “heads” or “tails.” Therefore $\Omega = \{\text{tails}, \text{heads}\}$ with $\lambda_1 = \text{tails}$ and $\lambda_2 = \text{heads}$. Moreover, if the coin is not weighted, it is reasonable to set $p_1 = 1/2 = p_2$. Indeed, if the experiment is repeated a great many times, the coin will just as often give tails as heads. A six-sided die, or any other game, can be treated in the same manner.

The idea of identifying the probability with the frequency of occurrence of a random event, which one might call the frequency interpretation of probability, is the one most commonly adopted by physicists. It is nevertheless interesting to consider the possibility that the notion of probability might not be identified with the frequency of occurrence. Indeed, as explained above, one may be led to consider as random a quantity of perfectly deterministic nature. In this case, it will not be possible to carry out independent experiments and the probability will not be identifiable with a quantity resulting from an experiment. Much work has been devoted to this question, but we are not concerned here with such theoretical discussions. Let us simply note that everyday life is far from contradicting the former standpoint. Indeed it is common to hear such statements as: “This horse has a three in four chance

of beating that one.” The race itself is a single event and the probability of $3/4$ mentioned here can in no way correspond to a frequency of occurrence. But this quantity may nevertheless prove useful to a gambler.

The set in question may be infinite. Consider a monkey typing on the keyboard of a computer. We may choose as random events the various possible words, i.e., sequences of letters that are not separated by a space. The set Ω of possible words is clearly infinite. To see this, we may imagine that the monkey typing on the computer keyboard never once presses on the space bar. One might object, quite rightly, that the animal has a finite lifespan, so that the set Ω must also be finite. Rather than trying to refine the example by finding a way around this objection, let us just say that it may be simpler to choose an infinite size for Ω than to estimate the maximal possible size. What matters in the end is the quality of the results obtained with the chosen model and the simplicity of that model.

Generalizing a little further, it should be noted that the set Ω of possible events may not only be infinite; it may actually be uncountable. In other words, it may be that the elements of the set Ω cannot be put into a one-to-one correspondence with the positive integers. To see this, suppose that we choose at random a real number between 0 and 1. In this case, we may identify Ω with an interval $[0, 1]$, and this is indeed uncountable in the above sense. This is a classic problem in mathematics. Let us outline our approach when the given set is uncountable. We consider the set of all subsets of Ω and we associate with every subset $\omega \subseteq \Omega$ a positive number $P(\omega)$. We then apply Kolmogorov's axioms to equip Ω with a probability law. To do so, the following conditions must be satisfied:

- $P(\Omega) = 1$ and $P(\emptyset) = 0$ (where $\emptyset$ is the empty set),
- if the subsets $A_1, A_2, \dots, A_n, \dots$ are pairwise disjoint, so that no pair of sets contains common elements, we must have¹

$$P(A_1 \cup A_2 \cup \dots \cup A_n \cup \dots) = P(A_1) + P(A_2) + \dots + P(A_n) + \dots$$

We see in this framework that we no longer speak of the probability of an event, but rather the probability of a set of events. Since a set may comprise a single element, the Kolmogorov axiom includes the definition of the probability of a single event. However, in the case where Ω is uncountably infinite, the probability of any single event will generally be zero. We shall return to this point when studying random variables, for the practical consequences here are very important.

2.2 Random Variables

A random variable is defined as a variable whose value is determined by a random experiment. More precisely, we consider a set Ω of random events λ

¹ $A \cup B$ denotes the set theoretic union of the two sets A and B .

and we associate with each of these events λ a value X_λ . If the possible values of X_λ are real numbers, we speak of a real random variable, whereas if they are complex numbers, we have a complex random variable. In the rest of this chapter we shall be concerned mainly with real- or integer-valued random variables. In the latter case, X_λ will be a whole number. In order to define the probability of a random variable, we proceed in two stages. We consider first the case where Ω is countable. The uncountable case will then lead to the idea of probability density.

If Ω is countable, we can define $p_i = P(\lambda_i)$ with $\sum_{i=1}^{\infty} p_i = 1$. The latter is also written $\sum_{\lambda \in \Omega} P(\lambda) = 1$, which simply means that the sum of the probabilities $P(\lambda)$ of each element of Ω must equal 1. Let x be a possible value of X_λ . Then $P(x)$ denotes the probability that X_λ is equal to x . We obtain this value by summing the probabilities of all random events in Ω such that $X_\lambda = x$. In the game of heads or tails, we may associate the value 0 to tails and 1 to heads. We thereby construct an integer-valued random variable. The probability $P(0)$ is thus $1/2$, as is $P(1)$. For a game with a six-sided die, we would have $P(1) = P(2) = \dots = P(6) = 1/6$. If our die is such that the number 1 appears on one side, the number 2 on two sides, and the number 3 on three sides, we then set $P(1) = 1/6$, $P(2) = 1/3$, and $P(3) = 1/2$.

Letting D_x denote the set of possible values of X_λ , we see that we must have

$$\sum_{x \in D_x} P(x) = 1 .$$

Note in passing that, although X_λ is indeed a random variable, x itself is a parameter and hence a known quantity. This comment may appear a subtle theoretical distinction. However, a lack of understanding of this point could lead the reader into great difficulties later on.

We are now in a position to state some examples of well known and extremely useful probability laws. Bernoulli's law is perhaps one of the simplest. The random variable, also known as a Bernoulli variable, can take only the values 0 and 1. The probability that X_λ equals 1 is denoted q and the probability that X_λ equals 0 is thus $1 - q$. Hence, q is the only parameter of the Bernoulli law.

Poisson's law is also widely used. In this case the random variable X_λ can take any positive integer values. If $P(n)$ is the probability that X_λ is equal to n , Poisson's law is defined by

$$P(n) = e^{-\mu} \frac{\mu^n}{n!} ,$$

where μ is the single parameter determining the distribution and

$$n! = n \cdot (n-1) \cdot (n-2) \cdot \dots \cdot 2 \cdot 1 .$$

As we shall discover later, the Poisson law is a simple model which allows us to describe a great many physical phenomena.

The situation is less simple when Ω is uncountable. It leads to further mathematical complexity in probability theory, requiring the use of measure theory. However, for our present purposes, it would not be useful to go into a detailed presentation of this subject. We shall therefore sidestep this difficulty by working directly with random variables. When we need to refer to random events to illustrate some physical concept, it will suffice to restrict ourselves to the countable case. Let us consider a random variable X_λ defined on an uncountable set Ω . In this case, the range of values of X_λ usually constitutes a continuous set. We then speak of a continuous random variable, as opposed to a discrete random variable which takes values in a countable set. The probability that X_λ is equal to some given precise value x is generally zero, rendering this notion somewhat irrelevant. It is basically for this reason that it is useful to introduce a distribution function $F_X(x)$ which gives the probability that X_λ is smaller than x . Letting ω_x be the subset of Ω containing those elements λ such that $X_\lambda \leq x$, we then have $P(\omega_x) = F_X(x)$. We define the probability density function of the variable X_λ as the derivative² of $F_X(x)$:

$$P_X(x) = \frac{dF_X(x)}{dx}.$$

As $F_X(\infty) = 1$ and $F_X(x) = \int_{-\infty}^x P_X(y)dy$, we deduce that

$$\int_{-\infty}^{\infty} P_X(x)dx = 1.$$

We can give a simple interpretation of the probability density function. Suppose that the probability density is continuous at point x . The probability that X_λ lies between $x - dx/2$ and $x + dx/2$ is then of the order of $P_X(x)dx$ for small dx and this relation is an equality in the limit as dx tends to 0. We should therefore bear in mind the fact that only $F_X(x)$ actually represents a probability, and that the same cannot be said of $P_X(x)$. Table 2.1 gives some of the more commonly encountered probability density functions in physics.

The appearance of the derivative should be treated with some caution from the mathematical standpoint, and precise mathematical conditions must be formulated in order to apply it correctly. One practical solution for simplifying the formalism consists in treating the quantities $F_X(x)$ and $P_X(x)$ as mathematical distributions. Here again we choose not to labor the details on this technical aspect. It is nevertheless worth noting that the use of distributions provides a unified framework for both discrete and continuous variables. Indeed, consider the probability law p_n , where n is a natural number. This law can be written in the form of a probability density function:

$$P_X(x) = \sum_n p_n \delta(x - n),$$

² We shall often need to consider this derivative in the sense of distributions.

Table 2.1. A selection of probability density functions commonly occurring in physics. Note that the function $\Gamma(x)$ is defined for positive x by $\Gamma(x) = \int_0^\infty u^{x-1} e^{-u} du$

Name	Probability density	Parameters
Uniform	$1/(b-a)$ if $a \leq x \leq b$ and 0 otherwise	a and b
Gaussian	$\frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x-m)^2}{2\sigma^2}\right]$	m and σ
Exponential	$a \exp(-ax)$ if $x \geq 0$ and 0 otherwise	a
Gamma	$\frac{\beta^\alpha x^{\alpha-1}}{\Gamma(\alpha)} \exp(-\beta x)$ if $x \geq 0$ and 0 otherwise	α and β
Cauchy	$\frac{a}{\pi(a^2 + x^2)}$ with $a > 0$	a

where $\delta(x-n)$ is the Dirac distribution centered on n . The latter is defined by its action on continuous functions $f(x)$, viz.,

$$\int f(x) \delta(x-n) dx = f(n) .$$

We thus find

$$\int f(x) P_X(x) dx = \sum_n p_n f(n) .$$

2.3 Means and Moments

The moments of a probability law play an important role in physics. They are defined by

$$\langle X_\lambda^r \rangle = \int x^r P_X(x) dx .$$

Moments are generally only considered for positive integer values of r . However, we may equally well choose any positive real value of r . It should be emphasized that there is *a priori* no guarantee of the existence of the moment of order r . For example, although all moments with $r \geq 0$ exist for the Gaussian law, only those moments $\langle X_\lambda^r \rangle$ with $r \in [0, 1)$ exist for the Cauchy law. Indeed, in the latter case, the moment $\langle X_\lambda \rangle$ is not defined because

$$\int_{-\infty}^{\infty} \frac{ax}{\pi(a^2 + x^2)} dx$$

is not a convergent integral.

For integer-valued discrete random variables, we have

$$\langle X_\lambda^r \rangle = \sum_n n^r p_n .$$

The first moments play a special role, because they are very often considered in physical problems. For $r = 1$, we obtain the mean value of the random variable:

$$m_X = \langle X_\lambda \rangle = \int x P_X(x) dx .$$

We also consider the variance, which is the second central moment

$$\sigma_X^2 = \langle (X_\lambda - m_X)^2 \rangle = \int (x - m_X)^2 P_X(x) dx .$$

The physical interpretation is simple. σ_X , also known as the standard deviation, represents the width of the probability density function, whilst m_X is the value on which the probability density is centered (see Fig. 2.2). Quite generally, we define the central moment of order r by $\langle (x - \langle x \rangle)^r \rangle$. Table 2.2 shows the mean and variance of various probability laws.

Table 2.2. Mean and variance of various probability laws

Name	Mean	Variance
Bernoulli	q	$q(1 - q)$
Poisson	μ	μ
Uniform	$(a + b)/2$	$(a - b)^2/12$
Gaussian	m	σ^2
Exponential	$1/a$	$1/a^2$
Gamma	α/β	α/β^2
Cauchy	Undefined	Undefined

The moments just defined correspond to mean values involving the probability density function. They are referred to as expected values or expectation values, or simply expectations. These quantities are defined from deterministic variables x rather than random variables X_λ . They are clearly deterministic quantities themselves. Although they characterize the probability density function, they cannot be directly ascertained by experiment. In fact, they can only be estimated. We shall return to the idea of estimation in Chapter 7. Let us content ourselves here with a property which illustrates the practical significance of moments. This property follows from the weak law of large

numbers, which tells us that if we consider independent realizations of random variables $Y_{\lambda(j)}$ distributed according to the same law with mean m_Y , then $S_\lambda(n)$ defined by

$$S_\lambda(n) = \frac{Y_{\lambda(1)} + Y_{\lambda(2)} + \cdots + Y_{\lambda(n)}}{n}$$

converges in probability to m_Y . Convergence in probability means that, as n tends to infinity ($n \rightarrow \infty$), the probability that $S_\lambda(n)$ takes a value different from m_Y tends to zero. We see that, if $\langle f(X_\lambda) \rangle$ exists [with $\langle f(X_\lambda) \rangle = \int f(x)P_X(x) dx$], and if we make n measurements of X_λ , then

$$\frac{f(X_{\lambda(1)}) + f(X_{\lambda(2)}) + \cdots + f(X_{\lambda(n)})}{n}$$

will be an approximation for $\langle f(X_\lambda) \rangle$ which improves as the number of measurements n increases. It is this basic property which confers a genuine practical meaning upon the definitions of expectation values. We may thus envisage them as the result of calculating the mean over an infinite number of random experiments which are independent but arise from the same statistical ensemble with the same probability law. Indeed it is this fact which justifies the name of expectation value. Once again, the reader is encouraged to reflect carefully on the idea of expectation value. It is often a misunderstanding of this basic notion which leads to serious errors of reasoning.

2.4 Median and Mode of a Probability Distribution

We have just observed that the Cauchy distribution has no mean because the integral

$$\lim_{a \rightarrow -\infty} \lim_{b \rightarrow \infty} \int_a^b x P_X(x) dx$$

does not converge. Note, however, that it is symmetrical about the origin, so that the value $x = 0$ must play some special role. This is indeed what we conclude when we introduce the notions of mode and median for this probability law.

When it exists, the median x_M is defined as the value of x such that

$$\int_{-\infty}^{x_M} P_X(x) dx = \int_{x_M}^{\infty} P_X(x) dx .$$

It is easy to check that the median of the Cauchy distribution is 0. It is also a straightforward matter to show that the median of the Gaussian distribution is just its mean. This last result, whereby the median is equal to the mean, is true for all probability laws that possess a mean and that are symmetrical about it.

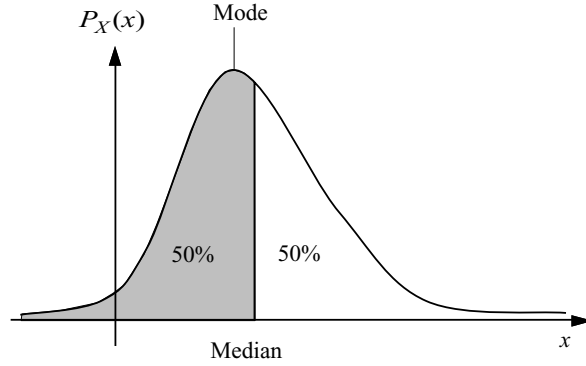


Fig. 2.1. Intuitive significance of the median and mode

The mode of a probability distribution corresponds to the most probable value in the case of discrete random variables and to the value at which the probability density function is maximal in the case of continuous random variables. The mode may not therefore consist of a unique value. Considering the case of continuous random variables, when the mode x_P exists and is unique, we must have

$$x_P = \operatorname{argmax}_x [P_X(x)] ,$$

which simply means that it is the value of x which maximizes $P_X(x)$, see Fig. 2.2. It is then clear that, if $P_X(x)$ is differentiable, the mode satisfies the relation

$$\frac{\partial}{\partial x} P_X(x_P) = 0 ,$$

or equivalently,

$$\frac{\partial}{\partial x} \ln [P_X(x_P)] = 0 .$$

It is thus easy to see that the mode of a Gaussian distribution is equal to its mean.

2.5 Joint Random Variables

It is common in physics to measure several quantities during an experiment. These quantities are not necessarily independent and we are often led to characterize their dependence. The notion of correlation partly achieves this aim. Let us define the idea of noise in a rather pragmatic way for the time being as

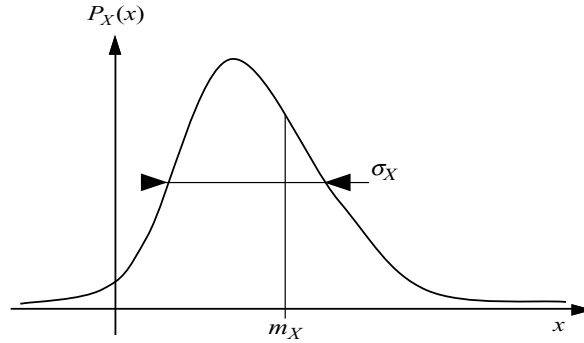


Fig. 2.2. Intuitive significance of mean and variance

something which impairs successive measurements of the same quantity with fluctuations. Imagine for example that we carry out a measurement using an experimental setup comprising a sensor and an amplifier. Suppose further that we wish to find out whether the noise observed during the recording of the measurements is noise arising in the sensor or noise arising in the amplifier. The notions of correlation and statistical dependence allow us to answer this practical question in a precise manner.

In a more mathematical context, suppose that for each random event λ we define two random variables X_λ and Y_λ . It may be interesting to consider the probability law for joint observation of certain values of X_λ and Y_λ . Likewise, knowing the value assumed by one of the random variables may provide information that will help us to determine the most likely values for the other random variable. Bayes' relation provides a precise formulation of this problem.

Consider two subsets A and B of Ω , where Ω represents the set of all possible random events. The probability of observing an event which belongs simultaneously to A and B is $P(A \cap B)$, where probabilities are measures of sets as illustrated in Fig. 2.3. $P(A \cup B)$ represents the probability that an element belongs to A or B . If we know *a priori* that the observed element belongs to B , the probability that it also belongs to A corresponds to the relative measure of A in B , that is, $P(A \cap B)/P(B)$, assuming of course that $P(B)$ is nonzero. If we use $P(A|B)$ to denote the probability that a random event belongs to A given that it belongs to B , we have Bayes' relation

$$P(A|B) = P(A, B) / P(B) ,$$

where we adopt the standard notation $P(A, B) = P(A \cap B)$. It follows immediately that

$$P(A, B) = P(A|B)P(B) = P(B|A)P(A) .$$

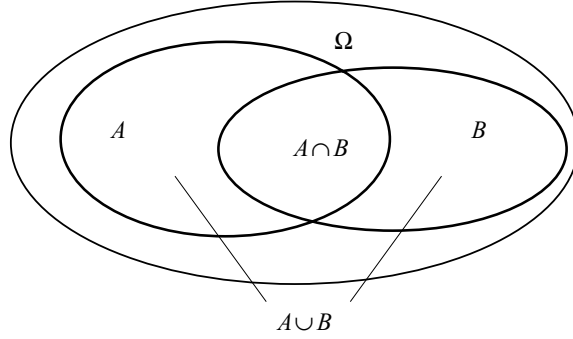


Fig. 2.3. Sets used in the definition of conditional probabilities

The joint distribution function $F_{X,Y}(x, y)$ relative to the two random variables X_λ and Y_λ is defined as the probability that simultaneously X_λ is less than x and Y_λ less than y . Let 1_x denote the set of random events in Ω such that X_λ is less than x . Likewise, let 1_y denote the set of random events in Ω such that Y_λ is less than y . Then it follows that $F_X(x) = P(1_x)$ and $F_Y(y) = P(1_y)$, and also that

$$F_{X,Y}(x, y) = P(1_x \cap 1_y) .$$

Since $P(1_x, 1_y) = P(1_x|1_y)P(1_y)$, we deduce that

$$F_{X,Y}(x, y) = F_{X|Y}(x|y)F_Y(y) .$$

We define the joint probability density function as

$$P_{X,Y}(x, y) = \frac{\partial^2 F_{X,Y}(x, y)}{\partial x \partial y} .$$

We have seen that $P_X(x) = \partial F_X(x)/\partial x$ and $P_Y(y) = \partial F_Y(y)/\partial y$, and we may now define the conditional probability density function as

$$P_{X|Y}(x|y) = \frac{P_{X,Y}(x, y)}{P_Y(y)} ,$$

and symmetrically,

$$P_{Y|X}(y|x) = \frac{P_{X,Y}(x, y)}{P_X(x)} .$$

To complete these relations, note that $F_{X,Y}(x, \infty) = F_X(x)$ and hence,

$$P_X(x) = \int_{-\infty}^{\infty} P_{X,Y}(x, y) dy ,$$

and likewise,

$$P_Y(y) = \int_{-\infty}^{\infty} P_{X,Y}(x, y) dx .$$

We say that $P_{X,Y}(x, y)$ is the joint probability law and that $P_X(x)$ and $P_Y(y)$ are the marginal probability laws.

2.6 Covariance

We are now in a position to give a precise definition of the independence of two random variables X_λ and Y_λ . These two random variables are independent if

$$P_{X,Y}(x, y) = P_X(x)P_Y(y) ,$$

which then implies that $P_{X|Y}(x|y) = P_X(x)$ and $P_{Y|X}(y|x) = P_Y(y)$. In other words, knowing the value of a realization of Y_λ tells us nothing about the value of X_λ since $P_{X|Y}(x|y) = P_X(x)$, and likewise, knowing the value of X_λ tells us nothing about the value of a realization of Y_λ .

The second extreme situation corresponds to the case where there is a perfectly deterministic relationship between X_λ and Y_λ , which we denote by $Y_\lambda = g(X_\lambda)$. Clearly, in this case, when the value of a realization of X_λ is known, only the value $g(X_\lambda)$ is possible for Y_λ , and we write

$$P_{Y|X}(y|x) = \delta(y - g(x)) .$$

Intermediate cases are interesting since they correspond to many practical situations. In order to measure the correlation between the two random variables X_λ and Y_λ , we might try to estimate the conditional probability density function $P_{X|Y}(x|y)$. However, the task is often impossible from a practical point of view and the notion of covariance is generally preferred. The covariance Γ_{XY} is defined by

$$\Gamma_{XY} = \langle X_\lambda Y_\lambda \rangle - \langle X_\lambda \rangle \langle Y_\lambda \rangle ,$$

or more explicitly,

$$\Gamma_{XY} = \iint (xy - m_X m_Y) P_{X,Y}(x, y) dx dy ,$$

where $m_X = \langle X_\lambda \rangle$ and $m_Y = \langle Y_\lambda \rangle$.

It can be shown that $|\Gamma_{XY}|^2 \leq \sigma_X^2 \sigma_Y^2$. Indeed, consider the quadratic form $(\alpha \delta X_\lambda - \delta Y_\lambda)^2$, where $\delta X_\lambda = X_\lambda - \langle X_\lambda \rangle$ and $\delta Y_\lambda = Y_\lambda - \langle Y_\lambda \rangle$. Since this form is positive definite, its expectation value must also be positive. Expanding out this expression, we obtain

$$\alpha^2 \langle (\delta X_\lambda)^2 \rangle - 2\alpha \langle \delta X_\lambda \delta Y_\lambda \rangle + \langle (\delta Y_\lambda)^2 \rangle \geq 0 .$$

The discriminant of this quadratic form in α must be negative, since it has at most one root. This implies that

$$\langle \delta X_\lambda \delta Y_\lambda \rangle^2 - \langle (\delta X_\lambda)^2 \rangle \langle (\delta Y_\lambda)^2 \rangle \leq 0 ,$$

which proves the claim.

It is therefore common to introduce the correlation coefficient, defined as $\rho_{XY} = \Gamma_{XY} / \sigma_X \sigma_Y$, which takes values between -1 and $+1$. From a practical standpoint, if the absolute value of ρ_{XY} is equal to 1, the two random variables are perfectly correlated. (To be precise, they must be proportional almost everywhere.) However, if ρ_{XY} is equal to 0, they are not correlated. This is the case, for example, if the two random variables X_λ and Y_λ are independent. It should nevertheless be borne in mind that, although the independence of two random variables does indeed imply that they are uncorrelated,³ i.e., that $\rho_{XY} = 0$, the converse is false. This property is easy to demonstrate and it is a straightforward matter to construct examples of dependent random variables for which $\rho_{XY} = 0$. Consider, for example, the random variable Φ_λ uniformly distributed between 0 and 2π . Then set $X_\lambda = \sin \Phi_\lambda$ and $Y_\lambda = \cos \Phi_\lambda$. It follows that $\langle X_\lambda Y_\lambda \rangle = \langle \sin \Phi_\lambda \cos \Phi_\lambda \rangle$ or

$$\langle X_\lambda Y_\lambda \rangle = \int \sin \phi \cos \phi P(\phi) d\phi = \frac{1}{2\pi} \int_0^{2\pi} \sin \phi \cos \phi d\phi = 0 .$$

The random variables X_λ and Y_λ are therefore uncorrelated. However, they are not independent, since $(X_\lambda)^2 + (Y_\lambda)^2 = 1$.

³ Introducing once again the centered variables $\delta X_\lambda = X_\lambda - \langle X_\lambda \rangle$ and $\delta Y_\lambda = Y_\lambda - \langle Y_\lambda \rangle$, it is easy to see that $\langle \delta X_\lambda \rangle = 0 = \langle \delta Y_\lambda \rangle$ and that $\Gamma_{XY} = \langle \delta X_\lambda \delta Y_\lambda \rangle$, i.e.,

$$\Gamma_{XY} = \iint xy P_{\delta X, \delta Y}(x, y) dx dy ,$$

noting that we are considering the probability density functions of the centered variables δX_λ and δY_λ . Clearly, we have $P_{\delta X}(x) = P_X(x - \langle X_\lambda \rangle)$ and $P_{\delta Y}(y) = P_Y(y - \langle Y_\lambda \rangle)$. Since by hypothesis $P_{X,Y}(x, y) = P_X(x)P_Y(y)$, we can deduce from the above that $P_{\delta X, \delta Y}(x, y) = P_{\delta X}(x)P_{\delta Y}(y)$. It thus follows that $\Gamma_{XY} = \iint xy P_{\delta X}(x)P_{\delta Y}(y) dx dy$ and hence

$$\Gamma_{XY} = \int x P_{\delta X}(x) dx \int y P_{\delta Y}(y) dy .$$

This in turn means that $\Gamma_{XY} = \langle \delta X_\lambda \rangle \langle \delta Y_\lambda \rangle$ and thus $\Gamma_{XY} = 0$.

2.7 Change of Variables

Given the probability density function $P_X(x)$ of a random variable X_λ , one often seeks in physics to determine the density of a related random variable $Y_\lambda = g(X_\lambda)$, where g is a function, assumed continuous. For example, in electromagnetism or optics, given the probability density function of the amplitude A_λ of the field, one may need to know the probability density function of the intensity $I_\lambda = |A_\lambda|^2$. In electronics, the output voltage V_λ of a component may depend on the applied voltage U_λ according to a relation of the form $V_\lambda = a \exp[\alpha(U_\lambda - U_0)]$. In order to determine the probability density function of fluctuations in the output physical quantity in terms of the probability density function of the input to the component, a change of variables calculation is required. This is the subject of the present section.

Suppose to begin with that the function $y = g(x)$ is increasing and differentiable, hence bijective. Let $F_X(x)$ and $F_Y(y)$ denote the distribution functions of X_λ and Y_λ . The probability that Y_λ is less than $g(x)$ is equal to the probability that X_λ is less than x . Hence, $F_Y[g(x)] = F_X(x)$. Differentiating, we obtain

$$\frac{dF_Y[g(x)]}{dx} = \frac{dF_Y[g(x)]}{dg(x)} \frac{dg(x)}{dx} = \frac{dF_X(x)}{dx}.$$

Moreover, since $y = g(x)$, writing

$$g'(x) = \frac{dg(x)}{dx},$$

we obtain

$$P_Y(y) = \frac{1}{g'(x)} P_X(x).$$

Noting that $g'(x) = dy/dx$, the above expression can also be written in the more memorable form (see Fig. 2.4)

$$P_Y(y)dy = P_X(x)dx.$$

If the relation $y = g(x)$ is not bijective, the above argument can be applied to intervals where it is bijective, adding the contributions from the various intervals for each value of y .

Considering the case where the probability density function $P_A(a)$ of the amplitude A_λ (assumed to be real-valued) of the electric field is Gaussian with zero mean and variance σ^2 , let us determine the probability density function $P_I(I)$ of the intensity $I_\lambda = |A_\lambda|^2$. To do so, we begin with the positive values of a . Hence,

$$P_I^+(I)dI = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{a^2}{2\sigma^2}\right) da.$$

Now $dI = 2a da$ and $a = \sqrt{I}$, which implies

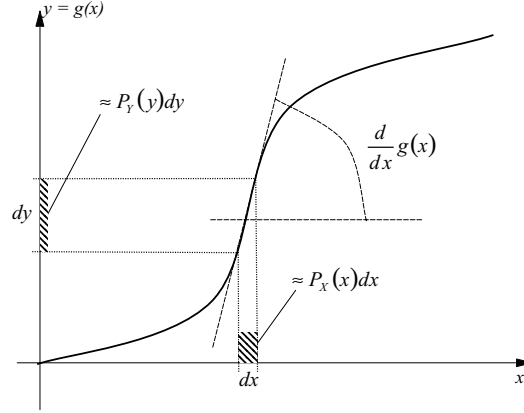


Fig. 2.4. Transformation of probability density upon change of variable

$$P_I^+(I) = \frac{1}{2\sqrt{2\pi I}\sigma} \exp\left(-\frac{I}{2\sigma^2}\right).$$

In the same manner we obtain for negative values

$$P_I^-(I) = \frac{1}{2\sqrt{2\pi I}\sigma} \exp\left(-\frac{I}{2\sigma^2}\right).$$

For each value of I , we may have $a = \sqrt{I}$ or $a = -\sqrt{I}$, and we thus deduce that $P_I(I) = P_I^+(I) + P_I^-(I)$. Hence,

$$P_I(I) = \frac{1}{\sqrt{2\pi I}\sigma} \exp\left(-\frac{I}{2\sigma^2}\right).$$

2.8 Stochastic Vectors

A stochastic vector $\mathbf{X}_\lambda$ is a vector whose value is determined from the outcome of a random experiment. As for random variables, we consider a set Ω of random events λ and associate a random vector $\mathbf{X}_\lambda$ with λ . If the possible values of the components of $\mathbf{X}_\lambda$ are real numbers, we shall speak of a real stochastic vector. If they are complex numbers, we have a complex stochastic vector.

For the moment, we discuss the case of real N -dimensional stochastic vectors. The stochastic vector can be described by its components, viz.,

$$\mathbf{X}_\lambda = (X_\lambda(1), X_\lambda(2), \dots, X_\lambda(N))^T,$$

where the symbol T indicates that we consider the transposed vector. We thus see that a stochastic vector is simply equivalent to a system of N random

variables. The distribution function $F_X(\mathbf{x})$ is the joint probability that $X_\lambda(j)$ is less than or equal to x_j for all j in the range from 1 to N , with $\mathbf{x} = (x_1, x_2, \dots, x_N)^T$. In other words,

$$F_X(\mathbf{x}) = \text{Prob}[X_\lambda(1) \leq x_1, X_\lambda(2) \leq x_2, \dots, X_\lambda(N) \leq x_N] .$$

In the case where the components are continuous variables, it is a simple matter to find the probability density function:

$$P_X(\mathbf{x}) = \frac{\partial^N}{\partial x_1 \partial x_2 \dots \partial x_N} F_X(\mathbf{x}) .$$

In the complex case, let $X_\lambda(j) = X_\lambda^R(j) + iX_\lambda^I(j)$, where $X_\lambda^R(j)$ and $X_\lambda^I(j)$ are the real and imaginary parts of the component $X_\lambda(j)$. The distribution function is then

$$F_X(\mathbf{x}) = \text{Prob}[X_\lambda^R(1) \leq x_1^R, X_\lambda^I(1) \leq x_1^I, X_\lambda^R(2) \leq x_2^R, X_\lambda^I(2) \leq x_2^I, \\ \dots, X_\lambda^R(N) \leq x_N^R, X_\lambda^I(N) \leq x_N^I] ,$$

and the probability density function is

$$P_X(\mathbf{x}) = \frac{\partial^{2N}}{\partial x_1^R \partial x_1^I \partial x_2^R \dots \partial x_N^R \partial x_N^I} F_X(\mathbf{x}) .$$

An N -dimensional complex stochastic vector is thus equivalent to a $2N$ -dimensional real stochastic vector.

The covariance matrix $\overline{\overline{I}}$ plays a central role in many situations. It is defined by its components:

$$\Gamma_{ij} = \langle X_\lambda(i) [X_\lambda(j)]^* \rangle - \langle X_\lambda(i) \rangle \langle [X_\lambda(j)]^* \rangle ,$$

where a^* is the complex conjugate of a . If the stochastic vector is real-valued, the above formula simplifies to

$$\Gamma_{ij} = \langle X_\lambda(i) X_\lambda(j) \rangle - \langle X_\lambda(i) \rangle \langle X_\lambda(j) \rangle .$$

The covariance matrix can be directly formulated in terms of the stochastic vector using

$$\overline{\overline{I}} = \langle \mathbf{X}_\lambda [\mathbf{X}_\lambda]^\dagger \rangle - \langle \mathbf{X}_\lambda \rangle \langle [\mathbf{X}_\lambda]^\dagger \rangle ,$$

where $\mathbf{a}^\dagger$ is the transposed conjugate of $\mathbf{a}$. Indeed, it is easy to see that

$$\Gamma_{ij} = \langle X_\lambda(i) [X_\lambda(j)]^* \rangle - \langle X_\lambda(i) \rangle \langle [X_\lambda(j)]^* \rangle$$

is equivalent to

$$\overline{\overline{I}} = \langle \mathbf{X}_\lambda [\mathbf{X}_\lambda]^\dagger \rangle - \langle \mathbf{X}_\lambda \rangle \langle [\mathbf{X}_\lambda]^\dagger \rangle .$$

Note that if $\mathbf{a}$ and $\mathbf{b}$ are two N -component vectors, $\mathbf{a}^\dagger \mathbf{b}$ is a scalar, since it is in fact the scalar product of $\mathbf{a}$ and $\mathbf{b}$, whilst $\mathbf{b} \mathbf{a}^\dagger$ is an $N \times N$ tensor with ij th component $b_i a_j^*$. This formulation is sometimes useful for simplifying certain proofs. For example, we can show that the covariance matrices are positive. For simplicity, we assume here that the mean value of $\mathbf{X}_\lambda$ is zero. If it is not, we can consider $\mathbf{Y}_\lambda = \delta \mathbf{X}_\lambda = \mathbf{X}_\lambda - \langle \mathbf{X}_\lambda \rangle$. For any vector $\mathbf{a}$, the modulus squared of the scalar product $\mathbf{a}^\dagger \mathbf{X}_\lambda$ is positive or zero, i.e.,

$$|\mathbf{a}^\dagger \mathbf{X}_\lambda|^2 \geq 0.$$

This expression can be written

$$(\mathbf{a}^\dagger \mathbf{X}_\lambda) [(\mathbf{X}_\lambda)^\dagger \mathbf{a}] \geq 0,$$

or

$$\mathbf{a}^\dagger \mathbf{X}_\lambda (\mathbf{X}_\lambda)^\dagger \mathbf{a} \geq 0.$$

Taking the expectation value of this expression, viz.,

$$\langle \mathbf{a}^\dagger \mathbf{X}_\lambda (\mathbf{X}_\lambda)^\dagger \mathbf{a} \rangle = \mathbf{a}^\dagger \langle \mathbf{X}_\lambda (\mathbf{X}_\lambda)^\dagger \rangle \mathbf{a},$$

and using the fact that

$$\overline{\overline{\mathbf{F}}} = \langle \mathbf{X}_\lambda [\mathbf{X}_\lambda]^\dagger \rangle,$$

we obtain for any $\mathbf{a}$ the relation

$$\mathbf{a}^\dagger \overline{\overline{\mathbf{F}}} \mathbf{a} \geq 0,$$

which shows that any covariance matrix is positive. From

$$\Gamma_{ij} = \langle X_\lambda(i) [X_\lambda(j)]^* \rangle,$$

we see immediately that we have a Hermitian matrix, i.e., $\overline{\overline{\mathbf{F}}}^\dagger = \overline{\overline{\mathbf{F}}}$, since

$$\{X_\lambda(j) [X_\lambda(i)]^*\}^* = X_\lambda(i) [X_\lambda(j)]^*.$$

Now it is well known that any Hermitian matrix can be diagonalized by a unitary transition matrix and that it has real eigenvalues. The covariance matrix is thus diagonalizable with zero or positive real eigenvalues and mutually orthogonal eigenvectors.

Now consider the example of real Gaussian N -dimensional stochastic vectors with mean $\mathbf{m}$ and covariance matrix $\overline{\overline{\mathbf{F}}}$. Let $\overline{\overline{\mathbf{K}}}$ be the inverse matrix of $\overline{\overline{\mathbf{F}}}$. Then the probability density function is

$$P_X(\mathbf{x}) = \frac{1}{(\sqrt{2\pi})^N \sqrt{|\overline{\overline{\mathbf{F}}}|}} \exp \left[-\frac{1}{2} (\mathbf{x} - \mathbf{m})^\dagger \overline{\overline{\mathbf{K}}} (\mathbf{x} - \mathbf{m}) \right],$$

where $|\overline{\overline{T}}|$ is the determinant of $\overline{\overline{T}}$. This expression can be written in the form

$$P_X(\mathbf{x}) = \frac{1}{(\sqrt{2\pi})^N \sqrt{|\overline{\overline{T}}|}} \exp \left[-\frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N (x_i - m_i) K_{ij} (x_j - m_j) \right] .$$

This simply means that

$$P_X(x_1, x_2, \dots, x_N) = \frac{1}{(\sqrt{2\pi})^N \sqrt{|\overline{\overline{T}}|}} \exp \left[-\frac{1}{2} Q(x_1, x_2, \dots, x_N) \right] ,$$

where

$$Q(x_1, x_2, \dots, x_N) = \sum_{i=1}^N \sum_{j=1}^N (x_i - m_i) K_{ij} (x_j - m_j) ,$$

and

$$\Gamma_{ij} = \int \dots \int (x_i x_j - m_i m_j) P_X(x_1, x_2, \dots, x_N) dx_1 dx_2 \dots dx_N .$$

Exercises

Exercise 2.1. Probability and Probability Density Function

Let X_λ be a random variable uniformly distributed between $-a$ and a , where $a > 0$. Consider the new variable Y_λ obtained from X_λ in the following way:

$$Y_\lambda = \begin{cases} -a/2 & \text{if } -a \leq X_\lambda \leq -a/2 , \\ X_\lambda & \text{if } -a/2 < X_\lambda < a/2 , \\ a/2 & \text{if } a/2 \leq X_\lambda \leq a . \end{cases}$$

Determine the probability density $P_Y(y)$ of Y_λ .

Exercise 2.2. Histogram Equalization

Let X_λ be a random variable with probability density function $P_X(x)$. Consider the new variable Y_λ obtained from X_λ in the following manner:

$$Y_\lambda = \int_{-\infty}^{X_\lambda} P_X(\eta) d\eta .$$

Determine the probability density function $P_Y(y)$ of Y_λ .

Exercise 2.3. Moments of the Gaussian Distribution

Calculate the central moments of the Gaussian probability law.

Exercise 2.4. Stochastic Vector

Consider a central Gaussian stochastic vector in two real dimensions. Show that we can write

$$P_{X_1, X_2}(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{(1-\rho^2)}} \exp \left[-\frac{x_1^2/\sigma_1^2 + x_2^2/\sigma_2^2 - 2x_1x_2\rho/\sigma_1\sigma_2}{2(1-\rho^2)} \right].$$

Exercise 2.5

Let $G(x, y)$ be the probability that the random variable X_λ lies between x and y . Determine the probability density of X_λ as a function of $G(x, y)$.

Exercise 2.6. Distribution of a Mixture

A gas contains a mixture of two types of atom A_1 and A_2 with respective concentrations c_1 and c_2 . The probability of photon emission by atoms A_1 is p_1 , whilst that for atoms A_2 is p_2 . What is the photon emission probability p for the mixed gas? Generalize to the case of an arbitrary mixture.

Exercise 2.7. Complex Gaussian Random Variable

Consider the complex random variable defined by $Z_\lambda = X_\lambda + iY_\lambda$ where $i^2 = -1$, and X_λ and Y_λ are independent Gaussian random variables with the same variance. Give an expression for the probability density of Z_λ .

Exercise 2.8. Weibull Variable

Determine the probability density function of Y_λ obtained from X_λ by the transformation $Y_\lambda = (X_\lambda)^\beta$, where $\beta > 0$ and X_λ is a random variable distributed according to the Gamma probability law. Analyze the special case where the Gamma distribution is exponential.

Exercise 2.9. Average of Noisy Measurements

A device measures a physical quantity g which is assumed to be constant in time. Several measurements with values F_i are made at N successive times. Each measurement is perturbed by noise B_i in such a way that

$$F_i = g + B_i.$$

B_i is a random variable, assumed to have a Gaussian distribution with mean 0 and standard deviation σ . For simplicity, assume that the dependence of the random variables on the random events is not noted. Assume also that the

variables B_i are statistically independent of one another. The sum of all the measurements is evaluated, thereby producing a new random variable

$$Y = \frac{1}{N} \sum_{i=1}^N F_i .$$

- (1) Calculate the probability density function of the random variable Y , assuming it obeys a Gaussian distribution.
- (2) Why can we say that measurement of g using Y is more ‘precise’ than measurement using a single value F_i ?

Exercise 2.10. Change of Variable

Consider two independent random variables X_λ and Y_λ , identically distributed according to a Gaussian probability law with zero mean. Determine the probability density function of the quotient random variable $Z_{\lambda_T} = X_{\lambda_1}/Y_{\lambda_2}$.

Noise Theory and Application to Physics
From Fluctuations to Information

Réfrégier, P.

2004, XIII, 288 p., Hardcover

ISBN: 978-0-387-20154-2