Chapter 2

Basic concepts of probability

Demetris Koutsoyiannis
Department of Water Resources and Environmental Engineering
Faculty of Civil Engineering, National Technical University of Athens, Greece

Summary

This chapter aims to serve as a reminder of basic concepts of probability theory, rather than a systematic and complete presentation of the theory. The text follows Kolmogorov's axiomatic foundation of probability and defines and discusses concepts such as random variables, distribution functions, independent and dependent events, conditional probability, expected values, moments and L moments, joint, marginal and conditional distributions, stochastic processes, stationarity, ergodicity, the central limit theorem, and the normal, χ^2 and Student distributions. Although the presentation is general and abstract, several examples with analytical and numerical calculations, as well as practical discussions are given, which focus on geophysical, and particularly hydrological, processes.

2.1 Axiomatic foundation of probability theory

For the understanding and the correct use of probability, it is very important to insist on the definitions and clarification of its fundamental concepts. Such concepts may differ from other, more familiar, arithmetic and mathematical concepts, and this may create confusion or even collapse of our cognitive construction, if we do not base it in concrete fundaments. For instance, in our everyday use of mathematics, we expect that all quantities are expressed by numbers and that the relationship between two quantities is expressed by the notion of a function, which to a numerical input quantity associates (maps) another numerical quantity, a unique output. Probability too does such a mapping, but the input quantity is not a number but an event, which mathematically can be represented as a set. Probability is then a quantified likelihood that the specific event will happen. This type of representation was proposed by Kolmogorov (1956)[*]. There are other probability systems different from Kolmogorov's axiomatic system, according to which the input is not a set. Thus, in Jaynes (2003)[†] the input of the mapping is a logical proposition and probability is a quantification of the plausibility of the proposition. The two systems are conceptually different but the differences mainly rely on

---

[*] Here we cite the English translation, second edition, whilst the original publication was in German in 1933.
[†] Jaynes's book that we cite here was published after his death in 1998.

interpretation rather than on the mathematical results. Here we will follow Kolmogorov's system.

Kolmogorov's approach to probability theory is based on the notion of *measure*, which maps *sets* onto *numbers*. The objects of probability theory, the *events*, to which probability is assigned, are thought of as sets. For instance the outcome of a roulette spin, i.e. the pocket in which the ball eventually falls on to the wheel is one of 37 (in a European roulette – 38 in an American one) pockets numbered 0 to 36 and coloured black or red (except 0 which is coloured green). Thus all sets {0}, {1}, … {36} are events (also called elementary events). But they are not the only ones. All possible subsets of Ω, including the empty set Ø, are events. The set $\Omega := \{0, 1, \dots, 36\}$ is an event too. Because any possible outcome is contained in Ω, the event Ω occurs in any case and it is called the *certain event*. The sets ODD := {1, 3, 5, …, 35}, EVEN := {2, 4, 6, …, 36}, RED := {1, 3, 5, 7, 9, 12, 14, 16, 18, 19, 21, 23, 25, 27, 30, 32, 34, 36}, and BLACK := Ω – RED – {0} are also events (in fact, betable). While events are represented as sets, in probability theory there are some differences from set theory in terminology and interpretation, which are shown in Table 2.1.

Table 2.1 Terminology correspondence in set theory and probability theory (adapted from Kolmogorov, 1956)

Set theory	Events
$A = \emptyset$	Event A is impossible
$A = \Omega$	Event A is certain
$AB = \emptyset$ (or $A \cap B = \emptyset$; disjoint sets)	Events A and B are incompatible (mutually exclusive)
$AB \dots N = \emptyset$	Events A, B, …, N are incompatible
$X := AB \dots N$	Event X is defined as the simultaneous occurrence of A, B, …, N
$X := A + B + \dots + N$ (or $X := A \cup B \cup \dots \cup N$)	Event X is defined as the occurrence of at least one of the events A, B, …, N
$X := A - B$	Event X is defined as the occurrence of A and, at the same time, the non-occurrence of B
$\overline{A}$ (the complementary of A)	The opposite event $\overline{A}$ consisting of the non-occurrence of A
$B \subset A$ (B is a subset of A)	From the occurrence of event B follows the inevitable occurrence of event A

Based on Kolmogorov's (1956) axiomatization, probability theory is based on three fundamental concepts and four axioms. The concepts are:

1. A non-empty set Ω, sometimes called the *basic set*, *sample space* or the *certain event* whose elements ω are known as *outcomes* or *states*.

2. A set Σ known as *σ-algebra* or *σ-field* whose elements E are subsets of Ω, known as *events*. Ω and Ø are both members of Σ, and, in addition, (a) if E is in Σ then the complement $\Omega - E$ is in Σ; (b) the union of countably many sets in Σ is also in Σ.
3. A function P called *probability* that maps events to real numbers, assigning each event E (member of Σ) a number between 0 and 1.

The triplet (Ω, Σ, P) is called *probability space*.

The four axioms, which define properties of P, are

Non-negativity: For any event A, $P(A) \geq 0$ (2.1.I)

Normalization: $P(\Omega) = 1$ (2.1.II)

Additivity: For any events A, B with $AB = \varnothing$, $P(A + B) = P(A) + P(B)$ (2.1.III)

IV. *Continuity at zero*: If $A_1 \supset A_2 \supset \ldots \supset A_n \supset \ldots$ is a decreasing sequence of events, with $A_1A_2\ldots A_n\ldots = \varnothing$, then $\lim_{n\to\infty}P(A_n) = 0$ (2.1.IV)

In the case that Σ is finite, axiom IV follows from axioms I-III; in the general case, however, it should be put as an independent axiom.

2.2 Random variables

A random variable X is a function that maps outcomes to numbers, i.e. quantifies the sample space Ω. More formally, a real single-valued function $X(\omega)$, defined on the basic set Ω, is called a *random variable* if for each choice of a real number a the set $\{X < a\}$ for all ω for which the inequality $X(\omega) < \alpha$ holds true, belongs to Σ.

With the notion of the random variable we can conveniently express events using basic mathematics. In most cases this is done almost automatically. For instance in the roulette case a random variable X that takes values 0 to 36 is intuitively assumed when we deal with a roulette experiment.

We must be attentive that a random variable is not a number but a function. Intuitively, we could think of a random variable as an object that represents simultaneously all possible states and only them. A particular value that a random variable may take in a random experiment, else known as a *realization* of the variable is a number. Usually we denote a random variable by an upper case letter, e.g. X, and its realization by a lower case letter, e.g. x. The two should not be confused. For example, if X represents the rainfall depth expressed in millimetres for a given rainfall episode (in this case Ω is the set of all possible rainfall depths) then $\{X \leq 1\}$ represents an *event* in the probability notion (a subset of Ω and a member of Σ – not to be confused with a physical event or episode) and has a probability $P\{X \leq 1\}$.* If x is a realization of X then $x \leq 1$ is not an event but a relationship between the two numbers x and 1,

---

* The consistent notation here would be $P(\{X \leq 1\})$. However, we simplified it dropping the parentheses; we will follow this simplification throughout this text. Some texts follow another convention, i.e., they drop the curly brackets.

which can be either true or false. In this respect it has no meaning to write $P\{x \le 1\}$. Furthermore, if we consider the two variables X and Y it is meaningful to write $P\{X \ge Y\}$ (i.e. $\{X \ge Y\}$ represents an event) but there is no meaning in the expression $P\{x \ge y\}$.

2.3 Distribution function

Distribution function is a function of the real variable x defined by

$$F_X(x) := P\{X \le x\} \tag{2.2}$$

where X is a random variable*. Clearly, $F_X(x)$ maps numbers (x values) to numbers. The random variable to which this function refers (is associated) is not an argument of the function; it is usually denoted as a subscript of F (or even omitted if there is no risk of confusion). Typically $F_X(x)$ has some mathematical expression depending on some parameters β_i. The domain of $F_X(x)$ is not identical to the range of the random variable X; rather it is always the set of real numbers. The distribution function is a non-decreasing function obeying the relationship

$$0 = F_X(-\infty) \le F_X(x) \le F_X(+\infty) = 1 \tag{2.3}$$

For its non-decreasing attitude, in the English literature the distribution function is also known as *cumulative distribution function* (cdf) – though cumulative is not necessary here. In hydrological applications the distribution function is also known as non-exceedence probability. Correspondingly, the quantity

$$F_X^*(x) := P\{X > x\} = 1 - F_X(x) \tag{2.4}$$

is known as exceedence probability, is a non-increasing function and obeys

$$1 = F_X^*(-\infty) \ge F_X^*(x) \ge F_X^*(+\infty) = 0 \tag{2.5}$$

The distribution function is always continuous on the right; however, if the basic set Ω is finite or countable, $F_X(x)$ is discontinuous on the left at all points x_i that correspond to outcomes ω_i, and it is constant in between consecutive points. In other words, the distribution function in these cases is staircase-like and the random variable is called *discrete*. If $F_X(x)$ is continuous, then the random variable is called *continuous*. A *mixed* case with a continuous part and a discrete part is also possible. In this case the distribution function has some discontinuities on the left, without being staircase-like.

The derivative of the distribution function

$$f_X(x) := \frac{dF(x)}{dx} \tag{2.6}$$

* In original Kolmogorov's writing $F_X(x)$ is defined as $P\{X < x\}$; however replacing '<' with '≤' makes the handling of distribution function more convenient and has prevailed in later literature.

is called the *probability density function* (sometimes abbreviated as pdf). In continuous variables, this function is defined everywhere but this is not the case in discrete variables, unless we use Dirac's δ functions. The basic properties of $f_X(x)$ are

$$f_X(x) \geq 0, \quad \int_{-\infty}^{\infty} f_X(x)dx = 1 \tag{2.7}$$

Obviously, the probability density function does not represent a probability; therefore it can take values higher than 1. Its relationship with probability is described by the following equation:

$$f_X(x) = \lim_{\Delta x \to 0} \frac{P\{x \leq X \leq x + \Delta x\}}{\Delta x} \tag{2.8}$$

The distribution function can be calculated from the density function by the following relationship, inverse of (2.6)

$$F_X(x) = \int_{-\infty}^{x} f_X(\xi)d\xi \tag{2.9}$$

For continuous random variables, the inverse function F_X^{-1} of $F_X(x)$ exists. Consequently, the equation $u = F_X(x)$ has a unique solution for x, that is $x_u = F_X^{-1}(u)$. The value x_u, which corresponds to a specific value u of the distribution function, is called *u-quantile* of the variable X.

2.3.1 An example of the basic concepts of probability

For clarification of the basic concepts of probability theory, we give the following example from hydrology. We are interested on the mathematical description of the possibilities that a certain day in a specific place and time of the year is wet or dry. These are the outcomes or states of our problem, so the basic set or sample space is

$$\Omega = \{\text{wet, dry}\}$$

The field Σ contains all possible events, i.e.,

$$\Sigma = \{\varnothing, \{\text{wet}\}, \{\text{dry}\}, \Omega\}$$

To fully define probability on Σ it suffices to define the probability of one of either states, say P(wet). In fact this is not easy – usually it is done by induction, and it needs a set of observations to be available and concepts of the *statistics* theory (see chapter 3) to be applied. For the time being let us arbitrarily assume that $P\{\text{wet}\} = 0.2$. The remaining probabilities are obtained by applying the axioms. Clearly, $P(\Omega) = 1$ and $P(\varnothing) = 0$. Since "wet" and "dry" are incompatible, $P\{\text{wet}\} + P\{\text{dry}\} = P(\{\text{wet}\} + \{\text{dry}\}) = P(\Omega) = 1$, so $P\{\text{dry}\} = 0.8$.

We define a random variable X based on the rule

$$X(\text{dry}) = 0, \quad X(\text{wet}) = 1$$

We can now easily determine the distribution function of X. For any $x < 0$,

$$F_X(x) = P\{X \leq x\} = 0$$

(because X, cannot take negative values). For $0 \leq x < 1$,

$$F_X(x) = P\{X \leq x\} = P\{X = 0\} = 0.8$$

Finally, for $x \geq 1$,

$$F_X(x) = P\{X \leq x\} = P\{X = 0\} + P\{X = 1\} = 1$$

The graphical depiction of the distribution function is shown on Fig. 2.1. The staircase-like shape reflects the fact that random variable is discrete.

If this mathematical model is to represent a physical phenomenon, we must have in mind that all probabilities depend on a specific location and a specific time of the year. So the model cannot be a global representation of the wet and dry state of a day. The model as formulated here is extremely simplified, because it does not make any reference to the succession of dry or wet states in different days. This is not an error; it simply diminishes the predictive capacity of the model. A better model would describe separately the probability of a wet day following a wet day, a wet day following a dry day (we anticipate that the latter should be smaller than the former), etc. We will discuss this case in section 2.4.2.

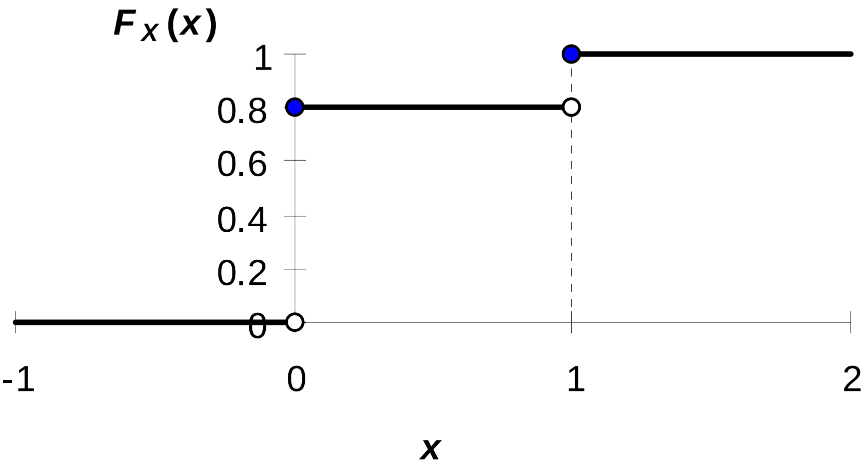

Fig. 2.1 Distribution function of a random variable representing the dry or wet state of a given day at a certain area and time of the year.

2.4 Independent and dependent events, conditional probability

Two events A and B are called *independent* (or *stochastically independent*), if

$$P(AB) = P(A)P(B) \tag{2.10}$$

Otherwise A and B are called (*stochastically*) *dependent*. The definition can be extended to many events. Thus, the events A_1, A_2, …, are *independent* if

$$P(A_{i_1} A_{i_2} \cdots A_{i_n}) = P(A_{i_1}) P(A_{i_2}) \cdots P(A_{i_n}) \tag{2.11}$$

for any finite set of distinct indices $i_1, i_2, \ldots, i_n$.

The handling of probabilities of independent events is thus easy. However, this is a special case because usually natural events are dependent. In the handling of dependent events the notion of *conditional probability* is vital. By definition (Kolmogorov, 1956), conditional probability of the event A given B (i.e. under the condition that the event B has occurred) is the quotient

$$P(A \mid B) := \frac{P(AB)}{P(B)} \tag{2.12}$$

Obviously, if $P(B) = 0$, this conditional probability cannot be defined, while for independent A and B, $P(A|B) = P(A)$. From (2.12) it follows that

$$P(AB) = P(A \mid B)P(B) = P(B \mid A)P(A) \tag{2.13}$$

and

$$P(B \mid A) := P(B)\frac{P(A \mid B)}{P(A)} \tag{2.14}$$

The latter equation is known as the *Bayes theorem*. It is easy to prove that the generalization of (2.11) for dependent events takes the forms

$$P(A_n \cdots A_1) = P(A_n \mid A_{n-1} \cdots A_1) \cdots P(A_2 \mid A_1)P(A_1) \tag{2.15}$$

$$P(A_n \cdots A_1 \mid B) = P(A_n \mid A_{n-1} \cdots A_1 B) \cdots P(A_2 \mid A_1 B)P(A_1 \mid B) \tag{2.16}$$

which are known as the *chain rules*. It is also easy to prove (homework) that if A and B are mutually exclusive, then

$$P(A + B \mid C) = P(A \mid C) + P(B \mid C) \tag{2.17}$$

$$P(C \mid A + B) = \frac{P(C \mid A)P(A) + P(C \mid B)P(B)}{P(A) + P(B)} \tag{2.18}$$

2.4.1 Some examples on independent events

a. Based on the example of section 2.3.1, calculate the probability that two consecutive days are wet assuming that the events in the two days are independent.

Let A := {wet} the event that a day is wet and $\overline{A}$ = {dry} the complementary event that a day is dry. As in section 2.3.1 we assume that $p := P(A) = 0.2$ and $q := P(\overline{A}) = 0.8$. Since we are interested on two consecutive days, our basic set will be

$$\Omega = \left\{A_1 A_2, \overline{A}_1 A_2, A_1 \overline{A}_2, \overline{A}_1 \overline{A}_2\right\}$$

where indices 1 and 2 correspond to the first and second day, respectively. By the independence assumption, the required probability will be

$$P_1 := (A_1 A_2) = P(A_1)P(A_2) = p^2 = 0.04$$

For completeness we also calculate the probabilities of all other events, which are:

$$P(\overline{A}_1 A_2) = P(A_1 \overline{A}_2) = pq = 0.16, \quad P(\overline{A}_1 \overline{A}_2) = q^2 = 0.64$$

As anticipated, the sum of probabilities of all events is 1.

b. Calculate the probability that two consecutive days are wet if it is known that one day is wet.

Knowing that one day is wet means that the event $\overline{A}_1 \overline{A}_2$ should be excluded (has not occurred) or that the composite event $A_1 A_2 + \overline{A}_1 A_2 + A_1 \overline{A}_2$ has occurred. Thus, we seek the probability

$$P_2 := P(A_1 A_2 \mid A_1 A_2 + \overline{A}_1 A_2 + A_1 \overline{A}_2)$$

which according to the definition of conditional probability is

$$P_2 = \frac{P\left(A_1 A_2 \left(A_1 A_2 + \overline{A}_1 A_2 + A_1 \overline{A}_2\right)\right)}{P\left(A_1 A_2 + \overline{A}_1 A_2 + A_1 \overline{A}_2\right)}$$

Considering that all combinations of events are mutually exclusive, we obtain

$$P_2 = \frac{P(A_1 A_2)}{P(A_1 A_2) + P(\overline{A}_1 A_2) + P(A_1 \overline{A}_2)} = \frac{p^2}{p^2 + 2pq} = \frac{p}{p + 2q} = 0.111\ldots$$

c. Calculate the probability that two consecutive days are wet if it is known that the first day is wet

Even though it may seem that this question is identical to the previous one, in fact it is not. In the previous question we knew that one day is wet, without knowing which one exactly. Here we have additional information, that the wet day is the first one. This information alters the probabilities as we will verify immediately.

Now we know that the composite event $A_1 A_2 + A_1 \overline{A}_2$ has occurred (events $\overline{A}_1 A_2$ and $\overline{A}_1 \overline{A}_2$ should be excluded). Consequently, the probability sought is

$$P_3 := P(A_1 A_2 \mid A_1 A_2 + A_1 \overline{A}_2)$$

which according to the definition of conditional probability is

$$P_3 = \frac{P\left(A_1 A_2 \left(A_1 A_2 + A_1 \overline{A}_2\right)\right)}{P\left(A_1 A_2 + A_1 \overline{A}_2\right)}$$

or

$$P_3 = \frac{P(A_1 A_2)}{P(A_1 A_2) + P(A_1 \overline{A}_2)} = \frac{p^2}{p^2 + pq} = \frac{p}{p + q} = p = 0.2$$

It is not a surprise that this is precisely the probability that one day is wet, as in section 2.3.1.

With these examples we demonstrated two important thinks: (a) that the prior information we have in a problem may introduce dependences in events that are initially assumed independent, and, more generally, (b) that the probability is not an objective and invariant quantity, characteristic of physical reality, but a quantity that depends on our knowledge or information on the examined phenomenon. This should not seem strange as it is always the case in science. For instance the location and velocity of a moving particle are not absolute objective quantities; they depend on the observer's coordinate system. The dependence of probability on given information, or its "subjectivity" should not be taken as ambiguity; there was nothing ambiguous in calculating the above probabilities, based on the information given each time.

2.4.2 An example on dependent events

The independence assumption in problem 2.4.1a is obviously a poor representation of the physical reality. To make a more realistic model, let us assume that the probability of today being wet (A_2) or dry $\overline{A}_2$ depend on the state yesterday (A_1 or $\overline{A}_1$). It is reasonable to assume that the following inequalities hold:

$$P(A_2 \mid A_1) > P(A_2) = p\,,\; P(\overline{A}_2 \mid \overline{A}_1) > P(\overline{A}_2) = q$$

$$P(A_2 \mid \overline{A}_1) < P(A_2) = p\,,\; P(\overline{A}_2 \mid A_1) < P(\overline{A}_2) = q$$

The problem now is more complicated than before. Let us arbitrarily assume that

$$P(A_2 \mid A_1) = 0.40, \quad P(A_2 \mid \overline{A}_1) = 0.15$$

Since

$$P(A_2 \mid A_1) + P(\overline{A}_2 \mid A_1) = 1$$

we can calculate

$$P(\overline{A}_2 \mid A_1) = 1 - P(A_2 \mid A_1) = 0.60$$

Similarly,

$$P(\overline{A}_2 \mid \overline{A}_1) = 1 - P(A_2 \mid \overline{A}_1) = 0.85$$

As the event $A_1 + \overline{A}_1$ is certain (i.e. $P(A_1 + \overline{A}_1) = 1$) we can write

$$P(A_2) = P(A_2 \mid A_1 + \overline{A}_1)$$

and using (2.18) we obtain

$$P(A_2) = P(A_2 \mid A_1)P(A_1) + P(A_2 \mid \overline{A}_1)P(\overline{A}_1) \tag{2.19}$$

It is reasonable to assume that the unconditional probabilities do not change after one day, i.e. that $P(A_2) = P(A_1) = p$ and $P(\overline{A}_2) = P(\overline{A}_1) = q = 1 - p$. Thus, (2.19) becomes

$$p = 0.40\,p + 0.15\,(1 - p)$$

from which we find $p = 0.20$ and $q = 0.80$. (Here we have deliberately chosen the values of $P(A_2 \mid A_1)$ and $P(A_2 \mid \overline{A}_1)$ such as to find the same p and q as in 2.4.1a).

Now we can proceed to the calculation of the probability that both days are wet:

$$P(A_2 A_1) = P(A_2 \mid A_1) P(A_1) = 0.4 \times 0.2 = 0.08 > p^2 = 0.04$$

For completeness we also calculate the probabilities of all other events, which are:

$$P(A_2 \overline{A}_1) = P(A_2 \mid \overline{A}_1) P(\overline{A}_1) = 0.15 \times 0.80 = 0.12, \; P(\overline{A}_2 A_1) = P(\overline{A}_2 \mid A_1) P(A_1) = 0.60 \times 0.20 = 0.12$$

$$P(\overline{A}_2 \overline{A}_1) = P(\overline{A}_2 \mid \overline{A}_1) P(\overline{A}_1) = 0.85 \times 0.80 = 0.68 > q^2 = 0.64$$

Thus, the dependence resulted in higher probabilities of consecutive events that are alike. This corresponds to a general natural behaviour that is known as *persistence* (see also chapter 4).

2.5 Expected values and moments

If X is a continuous random variable and $g(X)$ is an arbitrary function of X, then we define as the *expected value* or *mean* of $g(X)$ the quantity

$$E[g(X)] := \int_{-\infty}^{\infty} g(x) f_X(x) dx \tag{2.20}$$

Correspondingly, for a discrete random variable X, taking on the values x_1, x_2, …,

$$E[g(X)] := \sum_{i=1}^{\infty} g(x_i) P(X = x_i) \tag{2.21}$$

For certain types of functions $g(X)$ we take very commonly used statistical parameters, as specified below:

1. For $g(X) = X^r$, where $r = 0, 1, 2, \ldots$, the quantity

$$m_X^{(r)} := E[X^r] \tag{2.22}$$

   is called the r*th moment* (or the r*th moment about the origin*) of X. For $r = 0$, obviously the moment is 1.

2. For $g(X) = X$, the quantity

$$m_X := E[X] \tag{2.23}$$

   (that is the first moment) is called the *mean* of X. An alternative, commonly used, symbol for $E[X]$ is μ_X.

3. For $g(X) = (X - m_X)^r$, where $r = 0, 1, 2, \ldots$, the quantity

$$\mu_X^{(r)} := E\left[(X - m_X)^r\right] \tag{2.24}$$

   is called the r*th central moment* of X. For $r = 0$ and 1 the central moments are respectively 1 and 0. The central moments are related to the moments about the origin by

$$\mu_X^{(r)} = m_X^{(r)} - \binom{r}{1} m_X^{(r-1)} m_X + \cdots + (-1)^j \binom{r}{j} m_X^{(r-j)} m_X^j + \cdots (-1)^r m_X^{(0)} m_X^r \tag{2.25}$$

These take the following forms for small r

$$\mu_X^{(2)} = m_X^{(2)} - m_X^2$$

$$\mu_X^{(3)} = m_X^{(3)} - 3m_X^{(2)} m_X + 2m_X^3 \tag{2.26}$$

$$\mu_X^{(4)} = m_X^{(4)} - 4m_X^{(3)} m_X + 6m_X^{(2)} m_X^2 - 3m_X^4$$

and can be inverted to read:

$$m_X^{(2)} = \sigma_X^2 + m_X^2$$

$$m_X^{(3)} = \mu_X^{(3)} + 3\sigma_X^2 m_X + m_X^3 \tag{2.27}$$

$$m_X^{(4)} = \mu_X^{(4)} + 4\mu_X^{(3)} m_X + 6\sigma_X^2 m_X^2 + m_X^4$$

4. For $g(X) = (X - m_X)^2$, the quantity

$$\sigma_X^2 := \mu_X^{(2)} = E\left[(X - m_X)^2\right] = E[X^2] - m_X^2 \tag{2.28}$$

   (that is the second central moment) is called the *variance* of X. The variance is also denoted as $\mathrm{Var}[X]$. Its square root, denoted as σ_X or StD[X] is called the standard deviation of X.

The above families of moments are the classical ones having been used for more than a century. More recently, other types of moments have been introduced and some of them have been already in wide use in hydrology. We will discuss two families.

5. For $g(X) = X\,[F(X)]^r$, where $r = 0, 1, 2, \ldots$, the quantity

$$\beta_X^{(r)} := E\{X\,[F(X)]^r\} = \int_{-\infty}^{\infty} x\,[F(x)]^r f(x)\,dx = \int_0^1 x(u)\,u^r\,du \tag{2.29}$$

   is called the r*th probability weighted moment* of X (Greenwood et al., 1979). All probability weighted moments have dimensions identical to those of X (this is not the case in the other moments described earlier).

6. For $g(X) = X\,P_{r-1}^*(F(X))$, where $r = 1, 2, \ldots$, $P_r^*(u)$ is the rth shifted Legendre polynomial, i.e.,

$$P_r^*(u) := \sum_{k=0}^{r} p_{r,k}^* u^k \text{ with } p_{r,k}^* := (-1)^{r-k}\binom{r}{k}\binom{r+k}{k} = \frac{(-1)^{r-k}(r+k)!}{(k!)^2(r-k)!}$$

   the quantity

$$\lambda_X^{(r)} := E[X P_{r-1}^*(F(X))] = \int_0^1 x(u)\;P_{r-1}^*(u)\,du \tag{2.30}$$

is called the r*th L moment* of X (Hosking, 1990). Similar to the probability weighted moments, the L moments have dimensions identical to those of X. The L moments are related to the probability weighted moments by

$$\lambda_X^{(r)} := \sum_{k=0}^{r-1} p_{r,k}^* \ \beta_X^{(r)} \tag{2.31}$$

which for the most commonly used r takes the specific forms

$$\begin{aligned} \lambda_X^{(1)} &= \beta_X^{(0)} \ (= m_X) \\ \lambda_X^{(2)} &= 2\ \beta_X^{(1)} - \beta_X^{(0)} \\ \lambda_X^{(3)} &= 6\beta_X^{(2)} - 6\beta_X^{(1)} + \beta_X^{(0)} \\ \lambda_X^{(4)} &= 20\beta_X^{(3)} - 30\beta_X^{(2)} + 12\ \beta_X^{(1)} - \beta_X^{(0)} \end{aligned} \tag{2.32}$$

In all above quantities the index X may be omitted if there is no risk of confusion. The first four moments, central moments and L moments are widely used in hydrological statistics as they have a conceptual or geometrical meaning easily comprehensible. Specifically, they describe the location, dispersion, skewness and kurtosis of the distribution as it is explained below. Alternatively, other statistical parameters with similar meaning are also used, which are also explained below.

2.5.1 Location parameters

Essentially, the mean describes the location of the centre of gravity of the shape defined by the probability density function and the horizontal axis (Fig. 2.2a). It is also equivalent with the static moment of this shape about the vertical axis (given that the area of the shape equals 1). Often, the following types of location parameters are also used:

1. The *mode*, or most probable value, x_p, is the value of x for which the density $f_X(x)$ becomes maximum, if the random variable is continuous, or, for discrete variables, the probability becomes maximum. If $f_X(x)$ has one, two or many maxima, we say that the distribution is unimodal, bi-modal or multi-modal, respectively.

2. The *median*, $x_{0.5}$, is the value for which $P\{X \le x_{0.5}\} = P\{X \ge x_{0.5}\} = 1/2$, if the random variable is continuous (analogously we can define it for a discrete variable). Thus, a vertical line at the median separates the shape of the density function in two equivalent parts each having an area of 1/2.

Generally, the mean, the mode and the median are not identical unless the density is has a symmetrical and unimodal shape.

2.5.2 Dispersion parameters

The variance of a random variable and its square root, the standard deviation, which has same dimensions as the random variable, describe a measure of the scatter or dispersion of the probability density around the mean. Thus, a small variance shows a concentrated distribution (Fig. 2.2b). The variance cannot be negative. The lowest possible value is zero and this corresponds to a variable that takes one value only (the mean) with absolute certainty. Geometrically it is equivalent with the moment of inertia about the vertical axis passing from the centre of gravity of the shape defined by the probability density function and the horizontal axis.

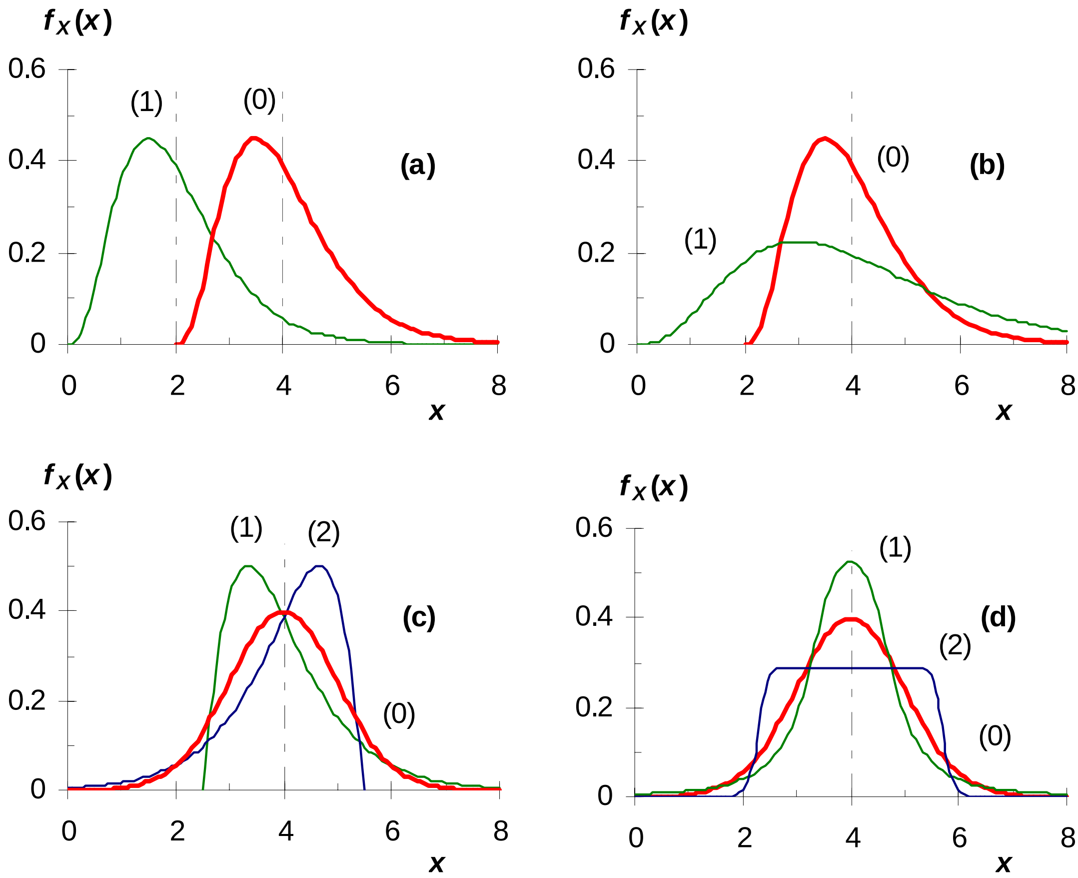

Fig. 2.2 Demonstration of the shape characteristics of the probability density function in relation to various parameters of the distribution function: *(a) Effect of the mean.* Curves (0) and (1) have means 4 and 2, respectively, whereas they both have standard deviation 1, coefficient of skewness 1 and coefficient of kurtosis 4.5. *(b) Effect of the standard deviation.* Curves (0) and (1) have standard deviation 1 and 2 respectively, whereas they both have mean 4, coefficient of skewness 1 and coefficient of kurtosis 4.5. *(c) Effect of the coefficient of skewness.* Curves (0), (1) and (2) have coefficients of skewness 0, +1.33 and -1.33, respectively, but they all have mean 4 and standard deviation 1; their coefficients of kurtosis are 3, 5.67 and 5.67, respectively. *(d) Effect of the coefficient of kurtosis.* Curves (0), (1) and (2) have coefficients of kurtosis 3, 5 and 2, respectively, whereas they all have mean 4, standard deviation 1 and coefficient of skewness 0.

Alternative measures of dispersion are provided by the so-called interquartile range, defined as the difference $x_{0.75} - x_{0.25}$, i.e. the difference of the 0.75 and 0.25 quantiles (or upper and lower quartiles) of the random variable (they define an area in the density function equal to 0.5), as well as the second L moment. This is well justified as it can be shown that the second L moment is the expected value of the difference between any two random realizations of the random variable.

If the random variable is positive, as happens with most hydrological variables, two dimensionless parameters are also used as measures of dispersion. These are called the *coefficient of variation* and the *L coefficient of variation*, and are defined, respectively, by:

$$C_{\mathrm{v}_X} := \frac{\sigma_X}{m_X}, \quad \tau_X^{(2)} := \frac{\lambda_X^{(2)}}{m_X} \tag{2.33}$$

2.5.3 Skewness parameters

The third central moment and the third L moment are used as measures of skewness. A zero value indicates that the density is symmetric. This can be easily verified from the definition of the third central moment. Furthermore, the third L moment indicates the expected value of the difference between the middle of three random realizations of a random variable from the average of the other two values (the smallest and the largest); more precisely the third central moment is the 2/3 of this expected value. Clearly then, in a symmetric distribution the distances of the middle value to the smallest and largest ones will be equal to each other and thus the third L moment will be zero. If the third central or L moment is positive or negative, we say that the distribution is positively or negatively skewed respectively (Fig. 2.2c). In a positively skewed unimodal distribution the following inequality holds: $x_{\mathrm{p}} \le x_{0.5} \le m_X$; the reverse holds for a negatively skewed distribution. More convenient measures of skewness are the following dimensionless parameters, named the *coefficient of skewness* and the *L coefficient of skewness*, respectively:

$$C_{\mathrm{s}_X} := \frac{\mu_X^{(3)}}{\sigma_X^3}, \quad \tau_X^{(3)} := \frac{\lambda_X^{(3)}}{\lambda_X^{(2)}} \tag{2.34}$$

2.5.4 Kurtosis parameters

The term kurtosis describes the "peakedness" of the probability density function around its mode. Quantification of this property provide the following dimensionless coefficients, based on the fourth central moment and the fourth L moment, respectively:

$$C_{\mathrm{k}_X} := \frac{\mu_X^{(4)}}{\sigma_X^4}, \quad \tau_X^{(4)} := \frac{\lambda_X^{(4)}}{\lambda_X^{(2)}} \tag{2.35}$$

These are called the c*oefficient of kurtosis* and the *L coefficient of kurtosis*. Reference values for kurtosis are provided by the normal distribution (see section 2.10.2), which has C_{k_X} = 3 and $\tau_X^{(4)}$ = 0.1226. Distributions with kurtosis greater than the reference values are called *leptokurtic* (acute, sharp) and have typically fat tails, so that more of the variance is due to infrequent extreme deviations, as opposed to frequent modestly-sized deviations. Distributions with kurtosis less than the reference values are called *platykurtic* (flat; Fig. 2.2d).

2.5.5 A simple example of a distribution function and its moments

We assume that the daily rainfall depth during the rain days, X, expressed in mm, for a certain location and time period, can be modelled by the *exponential distribution*, i.e.,

$$F_X(x) = 1 - e^{-x/\lambda}, \quad x \geq 0$$

where λ = 20 mm. We will calculate the location, dispersion, skewness and kurtosis parameters of the distribution.

Taking the derivative of the distribution function we calculate the probability density function:

$$f_X(x) = (1/\lambda)e^{-x/\lambda}, \quad x \geq 0$$

Both the distribution and the density functions are plotted in Fig. 2.3. To calculate the mean, we apply (2.20) for $g(X) = X$:

$$m_X = E[X] = \int_{-\infty}^{\infty} x f_X(x) dx = (1/\lambda)\int_0^{\infty} x e^{-x/\lambda} dx$$

After algebraic manipulations:

$$m_X = \lambda = 20\,\text{mm}$$

In a similar manner we find that for any $r \geq 0$

$$m_X^{(r)} = E[X^r] = r!\lambda^r$$

and finally, applying (2.26)

$$\sigma_X^2 = \lambda^2 = 400\,\text{mm}^2, \mu_X^{(3)} = 2\lambda^3 = 16000\,\text{mm}^3$$

$$\mu_X^{(4)} = 9\lambda^4 = 1440000\,\text{mm}^4$$

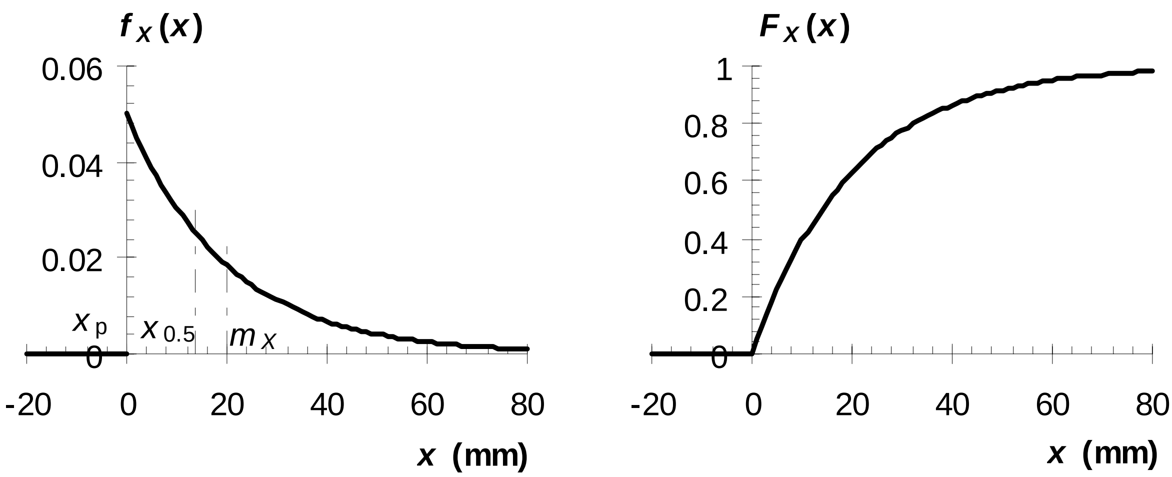

Fig. 2.3 Probability density function and probability distribution function of the exponential distribution, modelling the daily rainfall depth at a hypothetical site and time period.

The mode is apparently zero (see Fig. 2.3). The inverse of the distribution function is calculated as follows:

$$F_X(x_u) = u \rightarrow 1 - e^{-x_u/\lambda} = u \rightarrow x_u = -\lambda \ln(1-u)$$

Thus, the median is $x_{0.5} = -20 \times \ln 0.5 = 13.9$ mm. We verify that the inequality $x_p \le x_{0.5} \le m_X$, which characterizes positively skewed distributions, holds.

The standard deviation is σ_X = 20 mm and the coefficient of variation C_{v_X} = 1. This is a very high value indicating high dispersion.

The coefficient of skewness is calculated for (2.34):

$$C_{s_X} = 2\lambda^3 / \lambda^3 = 2$$

This verifies the positive skewness of the distribution, as also shown in Fig. 2.3. More specifically, we observe that the density function has an inverse-J shape, in contrast to other, more familiar densities (e.g. in Fig. 2.2) that have a bell-shape.

The coefficient of kurtosis is calculated from (2.35):

$$C_{k_X} = 9\lambda^4 / \lambda^4 = 9$$

Its high value shows that the distribution is leptokurtic, as also depicted in Fig. 2.3.

We proceed now in the calculations of probability weighted and L moments as well as other parameters based on these. From (2.29) we find

$$\beta_X^{(r)} = \int_0^1 x(u)\, u^r\, du = -\lambda \int_0^1 \ln(1-u)\, u^r\, du = \frac{\lambda}{r+1} \sum_{i=1}^{r+1} \frac{1}{i} \tag{2.36}$$

(This was somewhat tricky to calculate). This results in

$$\beta_X^{(0)} = \lambda, \;\; \beta_X^{(1)} = \frac{3\lambda}{4}, \;\; \beta_X^{(2)} = \frac{11\lambda}{18}, \;\; \beta_X^{(3)} = \frac{25\lambda}{48} \tag{2.37}$$

Then, from (2.32) we find the first four L moments and the three L moment dimensionless coefficients as follows:

$$\lambda_X^{(1)} = \lambda = 20 \text{ mm } (= m_X)$$

$$\lambda_X^{(2)} = 2\frac{3\lambda}{4} - \lambda = \frac{\lambda}{2} = 10 \text{ mm}$$

$$\lambda_X^{(3)} = 6\frac{11\lambda}{18} - 6\frac{3\lambda}{4} + \lambda = \frac{\lambda}{6} = 3.33 \text{ mm}$$

$$\lambda_X^{(4)} = 20\frac{25\lambda}{48} - 30\frac{11\lambda}{18} + 12\frac{3\lambda}{4} - \lambda = \frac{\lambda}{12} = 1.67 \text{ mm}$$

$$\tau_X^{(2)} = \frac{\lambda_X^{(2)}}{\lambda_X^{(1)}} = \frac{1}{2} = 0.5, \;\; \tau_X^{(3)} = \frac{\lambda_X^{(3)}}{\lambda_X^{(2)}} = \frac{1}{3} = 0.333, \;\; \tau_X^{(4)} = \frac{\lambda_X^{(4)}}{\lambda_X^{(2)}} = \frac{1}{6} = 0.\,167$$

Despite the very dissimilar values in comparison to those of classical moments, the results indicate the same behaviour, i.e., that the distribution is positively skewed and leptokurtic. In the following chapters we will utilize both classical and L moments in several hydrological problems.

2.5.6 Time scale and distribution shape

In the above example we saw that the distribution of a natural quantity such as rainfall, which is very random and simultaneously takes only nonnegative values, at a fine timescale, such as daily, exhibits high variation, strongly positive skewness and inverted-J shape of probability density function, which means that the most probable value (mode) is zero. Clearly, rainfall cannot be negative, so its distribution cannot be symmetric. It happens that the main body of rainfall values are close to zero, but a few values are extremely high (with low probability), which creates the distribution tail to the right. As we will see in other chapters, the distribution tails are even longer (or fatter, stronger, heavier) than described by this simple exponential distribution. In the exponential distribution, as demonstrated above, all moments (for any arbitrarily high but finite value of r) exist, i.e. take finite values. This is not, however, the case in long-tail distributions, whose moments above a certain rank r^* diverge, i.e. are infinite.

As we proceed from fine to coarser scales, e.g. from the daily toward the annual scale, aggregating more and more daily values, all moments increase but the standard deviation increases at a smaller rate in comparison to the mean, so the coefficient of variation decreases. In a similar manner, the coefficients of skewness and kurtosis decrease. Thus, the distributions tend to become more symmetric and the density functions take a more bell-shaped pattern. As we will se below, there are theoretical reasons for this behaviour for coarse timescales, which are related to the *central limit theorem* (see section 2.10.1). A more general theoretical explanation of the observed natural behaviours both in fine and coarse timescales is offered by the principle of *maximum entropy* (Koutsoyiannis, 2005a, b).

2.6 Change of variable

In hydrology we often prefer to use in our analyses, instead of the variable X that naturally describes a physical phenomenon (such as the rainfall depth in the example above), another variable Y which is a one-to-one mathematical transformation of X, e.g. $Y = g(X)$. If X is modelled as a random variable, then Y should be a random variable, too. The event $\{Y \le y\}$ is identical with the event $\{X \le g^{-1}(y)\}$ where g^{-1} is the inverse function of g. Consequently, the distribution functions of X and Y are related by

$$F_Y(y) = P\{Y \le y\} = P\{X \le g^{-1}(y)\} = F_X\left(g^{-1}(y)\right) \quad (2.38)$$

In the case that the variables are continuous and the function g differentiable, it can be shown that the density function of Y is given from that of X by

$$f_Y(y) = \frac{f_X\left(g^{-1}(y)\right)}{\left|g'\left(g^{-1}(y)\right)\right|} \tag{2.39}$$

where g' is the derivative of g. The application of (2.39) is elucidated in the following examples.

2.6.1 Example 1: the standardized variable

Very often the following transformation of a natural variable X is used:

$$Z = (X - m_X) / \sigma_X$$

This is called the *standardized variable*, is dimensionless and, as we will prove below, it has (a) zero mean, (b) unit standard deviation, and (c) third and fourth central moments equal to the coefficients of skewness and kurtosis of X, respectively.

From (2.38), setting $X = g^{-1}(Z) = \sigma_X Z + m_X$, we directly obtain

$$F_Z(z) = F_X\left(g^{-1}(z)\right) = F_X(\sigma_X z + m_X)$$

Given that $g'(x) = 1 / \sigma_X$, from (2.39) we obtain

$$f_Z(z) = \frac{f_X\left(g^{-1}(z)\right)}{\left|g'\left(g^{-1}(z)\right)\right|} = \sigma_X f_X(\sigma_X z + m_X)$$

Besides, from (2.20) we get

$$E[Z] = E[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x) dx = \int_{-\infty}^{\infty} \frac{x - m_X}{\sigma_X} f_X(x) dx =$$

$$= \frac{1}{\sigma_X} \int_{-\infty}^{\infty} x f_X(x) dx - \frac{m_X}{\sigma_X} \int_{-\infty}^{\infty} f_X(x) dx = \frac{1}{\sigma_X} m_X - \frac{m_X}{\sigma_X} 1$$

and finally

$$m_Z = E[Z] = 0$$

This entails that the moments about the origin and the central moments of Z are identical. Thus, the rth moment is

$$E\left[Z^r\right] = E\left[(g(X))^r\right] = \int_{-\infty}^{\infty} \left(\frac{x - m_X}{\sigma_X}\right)^r f_X(x) dx =$$

$$= \frac{1}{\sigma_X^r} \int_{-\infty}^{\infty} (x - m_X)^r f_X(x) dx = \frac{1}{\sigma_X^r} \mu_X^{(r)}$$

and finally

$$\mu_Z^{(r)} = m_Z^{(r)} = \frac{\mu_X^{(r)}}{\sigma_X^r}$$

2.6.2 Example 2: The exponential transformation and the Pareto distribution

Assuming that the variable X has exponential distribution as in the example of section 2.5.5, we will study the distribution of the transformed variable $Y = e^X$. The density and distribution of X are

$$f_X(x) = (1/\lambda)e^{-x/\lambda}, F_X(x) = 1 - e^{-x/\lambda}$$

and our transformation has the properties

$$Y = g(X) = e^X, g^{-1}(Y) = \ln Y, g'(X) = e^X$$

where $X \geq 0$ and $Y \geq 1$. From (2.38) we obtain

$$F_Y(y) = F_Y\left(g^{-1}(y)\right) = F_X(\ln y) = 1 - e^{-\ln y/\lambda} = 1 - y^{-1/\lambda}$$

and from (2.39)

$$f_Y(y) = \frac{f_X\left(g^{-1}(y)\right)}{\left|g'\left(g^{-1}(y)\right)\right|} = \frac{(1/\lambda)e^{-\ln y/\lambda}}{e^{\ln y}} = \frac{\lambda y^{-\lambda}}{y} = (1/\lambda)y^{-(1/\lambda+1)}$$

The latter can be more easily derived by taking the derivative of $F_Y(y)$.

This specific distribution is known as the *Pareto distribution*. The rth moment of this distribution is

$$m_Y^{(r)} = E\left[Y^r\right] = \int_{-\infty}^{\infty} y^r f_Y(y)dy = \int_1^{\infty} \lambda y^{r-1/\lambda-1}dx = \left.\frac{y^{r-1/\lambda}}{\lambda(r-1/\lambda)}\right]_{y=1}^{\infty} = \begin{cases} \dfrac{1}{1-r\lambda}, & r < 1/\lambda \\ \infty, & r \geq 1/\lambda \end{cases}$$

This clearly shows that only a finite number of moments ($r < 1/\lambda$) exist for this distribution, which means that the Pareto distribution has a long-tail.

2.7 Joint, marginal and conditional distributions

In the above sections, concepts of probability pertaining to the analysis of a single variable X have been described. Often, however, the simultaneous modelling of two (or more) variables is necessary. Let the couple of random variables (X, Y) represent two sample spaces (Ω_X, Ω_Y), respectively. The intersection of the two events $\{X \leq x\}$ and $\{Y \leq y\}$, denoted as $\{X \leq x\} \cap \{Y \leq y\} \equiv \{X \leq x, Y \leq y\}$ is an event of the sample space $\Omega_{XY} = \Omega_X \times \Omega_Y$. Based on the latter event, we can define the *joint probability distribution function* of (X, Y) as a function of the real variables (x, y):

$$F_{XY}(x, y) := P\{X \leq x, Y \leq y\} \tag{2.40}$$

The subscripts X, Y can be omitted if there is no risk of ambiguity. If F_{XY} is differentiable, then the function

$$f_{XY}(x, y) := \frac{\partial^2 F_{XY}(x, y)}{\partial x \partial y} \tag{2.41}$$

is the *joint probability density function* of the two variables. Obviously, the following equation holds:

$$F_{XY}(x,y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f_{XY}(\xi,\omega)d\omega d\xi \tag{2.42}$$

The functions

$$\begin{aligned} F_X(x) &:= P(X \le x) = \lim_{y\to\infty} F_{XY}(x,y) \\ F_Y(y) &:= P(Y \le y) = \lim_{x\to\infty} F_{XY}(x,y) \end{aligned} \tag{2.43}$$

are called the *marginal probability distribution functions* of X and Y, respectively. Also, the *marginal probability density functions* can be defined, from

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x,y)dy, \quad f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(x,y)dx \tag{2.44}$$

Of particular interest are the so-called *conditional probability distribution function* and *conditional probability density function* of X for a specified value of $Y = y$; these are given by

$$F_{X|Y}(x \mid y) = \frac{\int_{-\infty}^{x} f_{XY}(\xi,y)d\xi}{f_Y(y)}, \quad f_{X|Y}(x \mid y) = \frac{f_{XY}(x,y)}{f_Y(y)} \tag{2.45}$$

respectively. Switching X and Y we obtain the conditional functions of Y.

2.7.1 Expected values - moments

The expected value of any given function $g(X, Y)$ of the random variables (X, Y) is defined by

$$E[g(X,Y)] := \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y) f_{XY}(x,y)dydx \tag{2.46}$$

The quantity $E[X^p Y^q]$ is called $p + q$ moment of X and Y. Likewise, the quantity $E[(X - m_X)^p (Y - m_Y)^q]$ is called the $p + q$ central moment of X and Y. The most common of the latter case is the 1+1 moment, i.e.,

$$\sigma_{XY} := E[(X - m_X)(Y - m_Y)] = E[XY] - m_X m_Y \tag{2.47}$$

known as *covariance* of X and Y and also denoted as $\mathrm{Cov}[X,Y]$. Dividing this by the standard deviations σ_X and σ_Y we define the *correlation coefficient*

$$\rho_{XY} := \frac{\mathrm{Cov}[X,Y]}{\sqrt{\mathrm{Var}[X]\ \mathrm{Var}[Y]}} \equiv \frac{\sigma_{XY}}{\sigma_X \sigma_Y} \tag{2.48}$$

which is dimensionless with values $-1 \le \rho_{XY} \le 1$. As we will see later, this is an important parameter for the study of the correlation of two variables.

The *conditional expected value* of a function $g(X)$ for a specified value y of Y is defined by

$$E[g(X) \mid y] \equiv E[g(X) \mid Y = y] := \int_{-\infty}^{\infty} g(x) f_{X|Y}(x \mid y) dx \tag{2.49}$$

An important quantity of this type is the conditional expected value of X:

$$E[X \mid y] \equiv E[X \mid Y = y] := \int_{-\infty}^{\infty} x f_{X|Y}(x \mid y) dx \tag{2.50}$$

Likewise, the conditional expected value of Y is defined. The conditional variance of X for a given $Y = y$ is defined as

$$\mathrm{Var}[X \mid Y = y] := E\left[(X - E[X \mid Y = y])^2 \mid Y = y\right] = \int_{-\infty}^{\infty} (x - E[X \mid Y = y])^2 f_{X|Y}(x \mid y) dx \tag{2.51}$$

or

$$\mathrm{Var}[X \mid y] \equiv \mathrm{Var}[X \mid Y = y] := E\left[X^2 \mid Y = y\right] - (E[X \mid Y = y])^2 \tag{2.52}$$

Both $E[X \mid Y = y] \equiv E[X \mid y] =: \eta(y)$ and $\mathrm{Var}[X \mid Y = y] \equiv \mathrm{Var}[X \mid y] =: \upsilon(y)$ are functions of the real variable y, rather than constants. If we do not specify in the condition the value y of the random variable Y, then the quantities $E[X \mid Y] = \eta(Y)$ and $\mathrm{Var}[X \mid Y] = \upsilon(Y)$ become functions of the random variable Y. Hence, they are random variables themselves and they have their own expected values, i.e.,

$$E[E[X \mid Y]] = \int_{-\infty}^{\infty} E[X \mid y] f_Y(y) dy, \quad E[\mathrm{Var}[X \mid Y]] = \int_{-\infty}^{\infty} \mathrm{Var}[X \mid y] f_Y(y) dy \tag{2.53}$$

It is easily shown that $E[E[X \mid Y]] = E[X]$.

2.7.2 Independent variables

The random variables (X, Y) are called independent if for any couple of values (x, y) the following equation holds:

$$F_{XY}(x, y) = F_X(x) F_Y(y) \tag{2.54}$$

The following equation also holds:

$$f_{XY}(x, y) = f_X(x) f_Y(y) \tag{2.55}$$

and is equivalent with (2.54). The additional equations

$$\sigma_{XY} = 0 \leftrightarrow \rho_{XY} = 0 \leftrightarrow E[XY] = E[X]E[Y] \tag{2.56}$$

$$E[X \mid Y = x] = E[X], \quad E[Y \mid X = x] = E[Y] \tag{2.57}$$

are simple consequences of (2.54) but not sufficient conditions for the variable (X, Y) to be independent. Two variables (X, Y) for which (2.56) holds are called *uncorrelated*.

2.7.3 Sums of variables

A consequence of the definition of the expected value (equation (2.46)) is the relationship

$$E[c_1 g_1(X,Y) + c_2 g_2(X,Y)] = c_1 E[g_1(X,Y)] + c_2 E[g_2(X,Y)] \tag{2.58}$$

where c_1 and c_2 are any constant values whereas g_1 and g_2 are any functions. Apparently, this property can be extended to any number of functions g_i. Applying (2.58) for the sum of two variables we obtain

$$E[X + Y] = E[X] + E[Y] \tag{2.59}$$

Likewise,

$$E\left[(X - m_X + Y - m_Y)^2\right] = E\left[(X - m_X)^2\right] + E\left[(Y - m_Y)^2\right] + 2E[(X - m_X)(Y - m_Y)] \tag{2.60}$$

which results in

$$\mathrm{Var}[X + Y] = \mathrm{Var}[X] + \mathrm{Var}[Y] + 2\mathrm{Cov}[X, Y] \tag{2.61}$$

The probability distribution function of the sum $Z = X + Y$ is generally difficult to calculate. However, if X and Y are independent then it can be shown that

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(z - w) f_Y(w) dw \tag{2.62}$$

The latter integral is known as the *convolution integral* of $f_X(x)$ and $f_Y(y)$.

2.7.4 An example of correlation of two variables

We study a lake with an area of 10 km^2 lying on an impermeable subsurface. The inflow to the lake during the month of April, composed of rainfall and catchment runoff, is modelled as a random variable with mean 4.0×10^6 m^3 and standard deviation 1.5×10^6 m^3. The evaporation from the surface of the lake, which is the only outflow, is also modelled as a random variable with mean 90.0 mm and standard deviation 20.0 mm. Assuming that inflow and outflow are stochastically independent, we seek to find the statistical properties of the water level change in April as well as the correlation of this quantity with inflow and outflow.

Initially, we express the inflow in the same units as the outflow. To this aim we divide the inflow volume by the lake area, thus calculating the corresponding change in water level. The mean is 4.0×10^6 / 10.0×10^6 = 0.4 m = 400.0 mm and the standard deviation 1.5×10^6 / 10.0×10^6 = 0.15 m = 150.0 mm.

We denote by X and Y the inflow and outflow in April, respectively and by Z the water level change in the same month. Apparently,

$$Z = X - Y \tag{2.63}$$

We are given the quantities

$$m_X = E[X] = 400.0 \text{ mm}, \ \sigma_X = \sqrt{Var[X]} = 150.0 \text{ mm}$$

$$m_Y = E[Y] = 90.0 \text{ mm}, \ \sigma_Y = \sqrt{Var[Y]} = 20.0 \text{ mm}$$

and we have assumed that the two quantities are independent, so that their covariance Cov[X, Y] = 0 (see 2.56) and their correlation $\rho_{XY} = 0$.

Combining (2.63) and (2.58) we obtain

$$E[Z] = E[X - Y] = E[X] - E[Y] \rightarrow m_Z = m_X - m_Y \tag{2.64}$$

or m_Z = 310.0 mm. Subtracting (2.63) and (2.64) side by side we obtain

$$Z - m_z = (X - m_X) - (Y - m_Y) \tag{2.65}$$

and squaring both sides we find

$$(Z - m_z)^2 = (X - m_X)^2 + (Y - m_Y)^2 - 2(X - m_X)(Y - m_Y)$$

which, by taking expected values in both sides, results in the following equation (similar to (2.61) except in the sign of the last term)

$$\mathrm{Var}[Z] = \mathrm{Var}[X - Y] = \mathrm{Var}[X] + \mathrm{Var}[Y] - 2\mathrm{Cov}[X, Y] \tag{2.66}$$

Since Cov[X, Y] = 0, (2.66) gives

$$\sigma_Z^2 = 150.0^2 + 20.0^2 = 22900.0 \text{ mm}^2$$

and σ_Z = 151.3 mm.

Multiplying both sides of (2.65) by ($X - m_X$) and then taking expected values we find

$$E[(Z - m_z)(X - m_X)] = E[(X - m_X)^2] - E[(X - m_X)(Y - m_Y)]$$

or

$$\mathrm{Cov}[Z, X] = \mathrm{Var}[X] - \mathrm{Cov}[X, Y] \tag{2.67}$$

in which the last term is zero. Thus,

$$\sigma_{ZY} = \sigma_X^2 = 150.0^2 = 22500.0 \text{ mm}^2$$

Consequently, the correlation coefficient of X and Z is

$$\rho_{ZX} = \sigma_{ZX} / (\sigma_Z \sigma_X) = 22500.0 / (151.3 \times 150.0) = 0.991$$

Likewise,

$$\mathrm{Cov}[Z, Y] = \mathrm{Cov}[X, Y] - \mathrm{Var}[Y] \tag{2.68}$$

The first term of the right hand side is zero and thus

$$\sigma_{ZY} = -\sigma_Y^2 = -20.0^2 = -400.0 \text{ mm}^2$$

Consequently, the correlation coefficient of Y and Z is

$$\rho_{ZY} = \sigma_{ZY} / (\sigma_Z \sigma_Y) = -400.0 / (151.3 \times 20.0) = -0.132$$

The positive value of ρ_{ZX} manifests the fact that the water level increases with the increase of inflow (positive correlation of X and Z). Conversely, the negative correlation of Y and Z ($\rho_{ZY} < 0$) corresponds to the fact that the water level decreases with the increase of outflow.

The large, close to one, value of ρ_{ZX} in comparison to the much lower (in absolute value) value of ρ_{ZY} reflects the fact that in April the change of water level depends primarily on the inflow and secondarily on the outflow, given that the former is greater than the latter and also has greater variability (standard deviation).

2.7.5 An example of dependent discrete variables

Further to the example of section 2.4.2, we introduce the random variables X and Y to quantify the events (wet or dry day) of today and yesterday, respectively. Values of X or Y equal to 0 and 1 correspond to a day being dry and wet, respectively. We use the values of conditional probabilities (also called transition probabilities) of section 2.4.2, which with the current notation are:

$$\pi_{1|1} := P\{X = 1|Y = 1\} = 0.40, \pi_{0|1} := P\{X = 0|Y = 1\} = 0.60$$

$$\pi_{1|0} := P\{X = 1|Y = 0\} = 0.15, \pi_{0|1} := P\{X = 0|Y = 1\} = 0.85$$

The unconditional or marginal probabilities, as found in section 2.4.2, are

$$p_1 := P\{X = 1\} = 0.20, p_0 := P\{X = 0\} = 0.80$$

and the joint probabilities, again as found in section 2.4.2, are

$$p_{11} := P\{X = 1, Y = 1\} = 0.08, p_{01} := P\{X = 0, Y = 1\} = 0.12$$

$$p_{10} := P\{X = 1, Y = 0\} = 0.12, p_{00} := P\{X = 0, Y = 0\} = 0.68$$

It is reminded that the marginal probabilities of Y were assumed equal to those of X, which resulted in time symmetry ($p_{01} = p_{10}$). It can be easily shown (homework) that the conditional quantities $\pi_i|_j$ can be determined from the joint p_{ij} and vice versa, and the marginal quantities p_i can be determined for either of the two series. Thus, from the set of the ten above quantities only two are independent (e.g. $\pi_{1|1}$ and $\pi_{1|0}$) and all others can be calculated from these two.

The marginal moments of X and Y are

$$E[X] = E[Y] = 0\, p_0 + 1\, p_1 = p_1 = 0.20,\ \ E[X^2] = E[Y^2] = 0^2\, p_0 + 1^2\, p_1 = p_1 = 0.20$$

$$\mathrm{Var}[X] = E[X^2] - E[X]^2 = 0.2 - 0.2^2 = 0.16 = \mathrm{Var}[Y]$$

and the 1+1 joint moment is

$$E[XY] = 0 \times 0\, p_{00} + 0 \times 1\, p_{01} + 1 \times 0\, p_{10} + 1 \times 1\, p_{11} = p_{11} = 0.08$$

so that the covariance is

$$\sigma_{XY} \equiv \mathrm{Cov}[X, Y] = E[XY] - E[X]\, E[Y] = 0.08 - 0.2^2 = 0.04$$

and the correlation coefficient

$$\rho_{XY} := \frac{\mathrm{Cov}[X,Y]}{\sqrt{\mathrm{Var}[X]\ \mathrm{Var}[Y]}} \equiv \frac{0.04}{0.16} = 0.25$$

If we know that yesterday was a dry day, the moments for today are calculated from (2.49)-(2.52), replacing the integrals with sums and the conditional density $f_{X|Y}$ with the conditional probabilities $\pi_{i|j}$:

$$E[X|Y=0] = 0\ \pi_{0|0} + 1\ \pi_{1|0} = \pi_{1|0} = 0.15,\ \ E[X^2|Y=0] = 0^2\ \pi_{0|0} + 1^2\ \pi_{1|0} = \pi_{1|0} = 0.15$$

$$\mathrm{Var}[X|Y=0] = 0.15 - 0.15^2 = 0.128$$

Likewise,

$$E[X|Y=1] = 0\ \pi_{0|1} + 1\ \pi_{1|1} = \pi_{1|1} = 0.40,\ \ E[X^2|Y=1] = 0^2\ \pi_{0|1} + 1^2\ \pi_{1|1} = \pi_{1|1} = 0.40$$

$$\mathrm{Var}[X|Y=1] = 0.40 - 0.40^2 = 0.24$$

We observe that in the first case, $\mathrm{Var}[X|Y=0] < \mathrm{Var}[X]$. This can be interpreted as a decrease of uncertainty for the event of today, caused by the information that we have for yesterday. However, in the second case $\mathrm{Var}[X|Y=1] > \mathrm{Var}[X]$. Thus, the information that yesterday was wet, increases uncertainty for today. However, on the average the information about yesterday results in reduction of uncertainty. This can be expressed mathematically by $E[\mathrm{Var}[X|Y]]$ defined in (2.53), which is a weighted average of the two $\mathrm{Var}[X|Y=j]$:

$$E\{\mathrm{Var}[X|Y]\} := \mathrm{Var}[X|Y=0]\ p_0 + \mathrm{Var}[X|Y=1]\ p_1$$

This yields

$$E\{\mathrm{Var}[X|Y\} := 0.128 \times 0.8 + 0.24 \times 0.2 = 0.15 < 0.16 = \mathrm{Var}[X]$$

2.8 Many variables

All above theoretical analyses can be easily extended to more than two random variables. For instance, the distribution function of the n random variables $X_1, X_2, \ldots, X_n$ is

$$F_{X_1,\cdots,X_n}(x_1,\ldots,x_n) := P\{X_1 \le x_1,\ldots,X_n \le x_n\} \tag{2.69}$$

and is related to the n-dimensional probability density function by

$$F_{X_1,\cdots,X_n}(x_1,\ldots,x_n) = \int_{-\infty}^{x_1}\cdots\int_{-\infty}^{x_n} f_{X_1,\cdots,X_n}(\xi_1,\ldots,\xi_n)d\xi_n\cdots d\xi_1 \tag{2.70}$$

The variables $X_1, X_2, \ldots, X_n$ are independent if for any $x_1, x_2, \ldots, x_n$ the following holds true:

$$F_{X_1,\cdots,X_n}(x_1,\cdots,x_n) = F_{X_1}(x_1)\ldots F_{X_n}(x_n) \tag{2.71}$$

The expected values and moments are defined in a similar manner as in the case of two variables, and the property (2.58) is generalized for functions g_i of many variables.

2.9 The concept of a stochastic process

An arbitrarily (usually infinitely) large family of random variables $X(t)$ is called a *stochastic process* (Papoulis, 1991). To each one of them there corresponds an index t, which takes values from an *index set* T. Most often, the index set refers to time. The time t can be either *discrete* (when T is the set of integers) or continuous (when T is the set of real numbers); thus we have respectively a *discrete-time* or a *continuous-time* stochastic process. Each of the random variables $X(t)$ can be either discrete (e.g. the wet or dry state of a day) or continuous (e.g. the rainfall depth); thus we have respectively a *discrete-state* or a *continuous-state* stochastic process. Alternatively, a stochastic process may be denoted as X_t instead of $X(t)$; the notation X_t is more frequent for discrete-time processes. The index set can also be a vector space, rather than the real line or the set of integers; this is the case for instance when we assign a random variable (e.g. rainfall depth) to each geographical location (a two dimensional vector space) or to each location and time instance (a three-dimensional vector space). Stochastic processes with multidimensional index set are also known as random fields.

A realization $x(t)$ of a stochastic process $X(t)$, which is a regular (numerical) function of the time t, is known as a *sample function*. Typically, a realization is observed at countable time instances (not in continuous time, even in a continuous-time process). This sequence of observations is also called a *time series*. Clearly then, a time series is a sequence of numbers, whereas a stochastic process is a family of random variables. Unfortunately, a large literature body does not make this distinction and confuses stochastic processes with time series.

2.9.1 Distribution function

The distribution function of the random variable X_t, i.e.,

$$F(x;t) := P\{X(t) \le x\} \tag{2.72}$$

is called *first order distribution function* of the process. Likewise, the *second order distribution function* is

$$F(x_1, x_2; t_1, t_2) := P\{X(t_1) \le x_1, X(t_2) \le x_2\} \tag{2.73}$$

and the n*th order distribution function*

$$F(x_1, \ldots, x_n; t_1, \ldots, t_n) := P\{X(t_1) \le x_1, \ldots, X(t_n) \le x_n\} \tag{2.74}$$

A stochastic process is completely determined if we know the nth order distribution function for any n. The nth order probability density function of the process is derived by taking the derivatives of the distribution function with respect to all x_i.

2.9.2 Moments

The moments are defined in the same manner as in sections 2.5 and 2.7.1. Of particular interest are the following:

1. The *process mean*, i.e. the expected value of the variable $X(t)$:

$$m(t) := E[X(t)] = \int_{-\infty}^{\infty} x f(x;t) dt \tag{2.75}$$

2. The process *autocovariance*, i.e. the covariance of the random variables $X(t_1)$ and $X(t_2)$:

$$C(t_1, t_2) := \mathrm{Cov}[X(t_1), X(t_2)] = E[(X(t_1) - m(t_1))(X(t_2) - m(t_2))] \tag{2.76}$$

The process variance (the variance of the variable $X(t)$), is $\mathrm{Var}[X(t)] = C(t, t)$. Consequently, the process autocorrelation (the correlation coefficient of the random variables $X(t_1)$ and $X(t_2)$) is

$$\rho(t_1, t_2) := \frac{\mathrm{Cov}[X(t_1), X(t_2)]}{\sqrt{Var[X(t_1)]Var[X(t_2)]}} = \frac{C(t_1, t_2)}{\sqrt{C(t_1, t_1)C(t_2, t_2)}} \tag{2.77}$$

2.9.3 Stationarity

As implied by the above notation, in the general setting, the statistics of a stochastic process, such as the mean and autocovariance, depend on time and thus vary with time. However, the case where these statistical properties remain constant in time is most interesting. A process with this property is called *stationary* process. More precisely, a process is called strict-sense stationary if all its statistical properties are invariant to a shift of time origin. That is, the distribution function of any order of $X(t + \tau)$ is identical to that of $X(t)$. A process is called *wide-sense stationary* if its mean is constant and its autocovariance depends only on time differences, i.e.

$$E[X(t)] = \mu, \quad E[(X(\tau) - \mu)\,(X(t + \tau) - \mu)] = C(\tau) \tag{2.78}$$

A strict-sense stationary process is also wide-sense stationary but the inverse is not true.

A process that is not stationary is called nonstationary. In a nonstationary process one or more statistical properties depend on time. A typical case of a nonstationary process is a cumulative process whose mean is proportional to time. For instance, let us assume that the rainfall intensity $\Xi(t)$ at a geographical location and time of the year is a stationary process, with a mean μ. Let us further denote $X(t)$ the rainfall depth collected in a large container (a cumulative raingauge) at time t and assume that at the time origin, $t = 0$, the container is empty. It is easy then to understand that $E[X(t)] = \mu\, t$. Thus $X(t)$ is a nonstationary process.

We should stress that stationarity and nonstationarity are properties of a process, not of a sample function or time series. There is some confusion in the literature about this, as a lot of studies assume that a time series is stationary or not, or can reveal whether the process is stationary or not. As a general rule, to characterise a process nonstationary, it suffices to show that some statistical property is a *deterministic* function of time (as in the above example of the raingauge), but this cannot be straightforwardly inferred merely from a time series.

Stochastic processes describing periodic phenomena, such as those affected by the annual cycle of Earth, are clearly nonstationary. For instance, the daily temperature at a mid-latitude location could not be regarded as a stationary process. It is a special kind of a nonstationary process, as its properties depend on time on a periodical manner (are periodic functions of time). Such processes are called *cyclostationary* processes.

2.9.4 Ergodicity

The concept of *ergodicity* (from the Greek words *ergon* – work – and *odos* – path) is central for the problem of the determination of the distribution function of a process from a single sample function (time series) of the process. A stationary stochastic process is ergodic if any statistical property can be determined from a sample function. Given that in practice, the statistical properties are determined as time averages of time series, the above definition can be stated alternatively as: a stationary stochastic process is ergodic if time averages equal ensemble averages (i.e. expected values). For example, a stationary stochastic process is *mean ergodic* if

$$E[X(t)] = \lim_{N\to\infty} \frac{1}{N}\sum_{t=0}^{N} X(t) \qquad \text{(for a discrete time process)}$$

$$E[X(t)] = \lim_{T\to\infty} \frac{1}{T}\int_{0}^{T} X(t)dt \qquad \text{(for a continuous time process)} \tag{2.79}$$

The left-hand side in the above equations represents the ensemble average whereas the right-hand side represents the time average, for the limiting case of infinite time. Whilst the left-hand side is a parameter, rather than a random variable, the right-hand side is a random variable (as a sum or integral of random variables). The equating of a parameter with a random variable implies that the random variable has zero variance. This is precisely the condition that makes a process ergodic, a condition that does not hold true for every stochastic process.

2.10 The central limit theorem and some common distribution functions

The *central limit theorem* is one of the most important in probability theory. It concerns the limit distribution function of a sum of random variables – components, which, under some conditions but irrespectively of the distribution functions of the components, is always the same, the celebrated *normal distribution*. This is the most commonly used distribution in probability theory as well as in all scientific disciplines and can be derived not only as a consequence of the central limit theorem but also from the principle of maximum entropy, a very powerful physical and mathematical principle (Papoulis, 1990, p. 422-430).

In this section we will present the central limit theorem, the normal distribution, and some other distributions closely connected to the normal (χ^2 and Student). All these distributions are fundamental in statistics (chapter 3) and are commonly used for statistical estimation and prediction. Besides, the normal distribution has several applications in hydrological statistics, which will be discussed in chapters 5 and 6.

2.10.1 The central limit theorem and its importance

Let X_i (i = 1, …, n) be random variables and let $Z := X_1 + X_2 + \cdots + X_n$ be its sum with $E[Z] = m_Z$ and $\text{Var}[Z] = s_Z^2$. The central limit theorem says that the distribution of Z, under some general conditions (see below) has a specific limit as n tends to infinity, i.e.,

$$F_Z(z) \xrightarrow[n\to\infty]{} \int_{-\infty}^{z} \frac{1}{\sigma_Z\sqrt{2\pi}} e^{-\frac{(\zeta - m_Z)^2}{2\sigma_Z^2}} d\zeta \tag{2.80}$$

and in addition, if X_i are continuous variables, the density function of Z has also a limit,

$$f_Z(z) \xrightarrow[n\to\infty]{} \frac{1}{\sigma_Z\sqrt{2\pi}} e^{-\frac{(z - m_Z)^2}{2\sigma_Z^2}} \tag{2.81}$$

The distribution function in the right-hand side of (2.80) is called the *normal* (or *Gauss*) *distribution* and, likewise, the function in the right-hand side of (2.81) is called the *normal probability density function*.

In practice, the convergence for $n \to \infty$ can be regarded as an approximation if n is sufficiently large. How large should n be so that the approximation be satisfactory, depends on the (joint) distribution of the components X_i. In most practical application, a value n = 30 is regarded to be satisfactory (with the condition that X_i are independent and identically distributed). Fig. 2.4 gives a graphical demonstration of the central limit theorem based on an example. Starting from independent random variables X_i with exponential distribution, which is positively skewed, we have calculated (using (2.62)) and depicted the distribution of the sum of 2, 4, 8, 16 and 32 variables. If the distribution of X_i were symmetric, the convergence would be much faster.

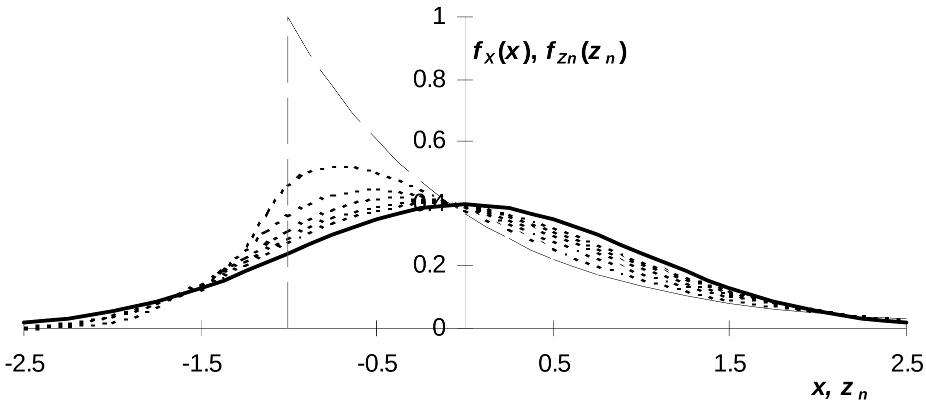

Fig. 2.4 Convergence of the sum of exponentially distributed random variables to the normal distribution (thick line). The dashed line with peak at $x = -1$ represents the probability density of the initial variables X_i, which is $f_X(x) = e^{-(x-1)}$ (mean 0, standard deviation 1). The dotted lines (going from the more peaked to the less peaked) represent the densities of the sums $Z_n = (X_1 + \ldots + X_n) / n$ for n = 2, 4, 8, 16 and 32. (The division of the sum by n helps for a better presentation of the curves, as all Z_i have the same mean and variance, 0 and 1, respectively, and does not affect the essentials of the central limit theorem.)

The conditions for the validity of the central limit theorem are general enough, so that they are met in many practical situations. Some sets of conditions (e.g. Papoulis, 1990, p. 215) with particular interest are the following: (a) the variables X_i are independent identically distributed with finite third moment; (b) the variables X_i are bounded from above and below with variance greater than zero; (c) the variables X_i are independent with finite third moment and the variance of Z tends to infinity as n tends to infinity. The theorem has been extended for variables X_i that are interdependent, but each one is effectively dependent on a finite number of other variables. Practically speaking, the central limit theorem gives satisfactory approximations for sums of variables unless the tail of the density functions of X_i are over-exponential (long, like in the Pareto example; see section 2.6.2) or the dependence of the variables is very strong and spans the entire sequence of X_i (long range dependence; see chapter 4). Note that the normal density function has an exponential tail (it can be approximated by an exponential decay for large x) and all its moments exist (are finite), whereas in over-exponential densities all moments beyond a certain order diverge. Since in hydrological processes the over-exponential tails, as well as the long-range dependence, are not uncommon, we must be attentive in the application of the theorem.

We observe in (2.80) and (2.81) that the limits of the functions $F_Z(z)$ and $f_Z(z)$ do not depend on the distribution functions of X_i, that is, the result is the same irrespectively of the distribution functions of X_i. Thus, provided that the conditions for the applicability of the theorem hold, (a) we can know the distribution function of the sum without knowing the distribution of the components, and (b) precisely the same distribution describes any variable that is a sum of a large number of components. Here lies the great importance of the normal distribution in all sciences (mathematical, physical, social, economical, etc.). Particularly, in statistics, as we will see in chapter 3, the central limit theorem implies that the sample average for any type of variables will have normal distribution (for a sample large enough).

In hydrological statistics, as we will see in chapters 5 and 6 in more detail, the normal distribution describes with satisfactory accuracy variables that refer to long time scales such as annual. Thus, the annual rainfall depth in a wet area is the sum of many (e.g. more than 30) independent rainfall events during the year (this, however, is not valid for rainfall in dry areas). Likewise, the annual runoff volume passing through a river section can be regarded as the sum of 365 daily volumes. These are not independent, but as an approximation, the central limit theorem can be applicable again.

2.10.2 The normal distribution

The random variable X is *normally distributed* or (Gauss distributed) with parameters μ and σ (symbolically $N(\mu, \sigma)$ if its probability density is

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{2.82}$$

The corresponding distribution function is

$$F_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{(\xi-\mu)^2}{2\sigma^2}} d\xi \quad (2.83)$$

The mean and standard deviation of X are μ and σ, respectively. The distribution is symmetric (Fig. 2.5) and thus its third central moment and its third L moment are zero. The fourth central moment is $3\sigma^4$ (hence $C_k = 3$) and the fourth L moment is $0.1226\lambda_X^{(2)}$ (hence $\tau_X^{(4)} = 0.1226$).

The integral in the right-hand side of (2.83) is not calculated analytically. Thus, the typical calculations ($x \to F_X(x)$ or $F_X(x) \to x$) are done either numerically or using tabulated values of the so-called *standard normal variate* Z, that is obtained from X with the transformation

$$Z = \frac{X - \mu}{\sigma} \leftrightarrow X = \mu + \sigma Z \quad (2.84)$$

and its distribution is $N(0,1)$. It is easy to obtain (see section 2.6.1) that

$$F_X(x) = F_Z(z) = F_Z\left(\frac{x-\mu}{\sigma}\right) \quad (2.85)$$

Such tables are included in all textbooks of probability and statistics, as well as in the Appendix of this text. However, nowadays all common numerical computer packages (including spreadsheet applications etc.) include functions for the direct calculation of the integral.*

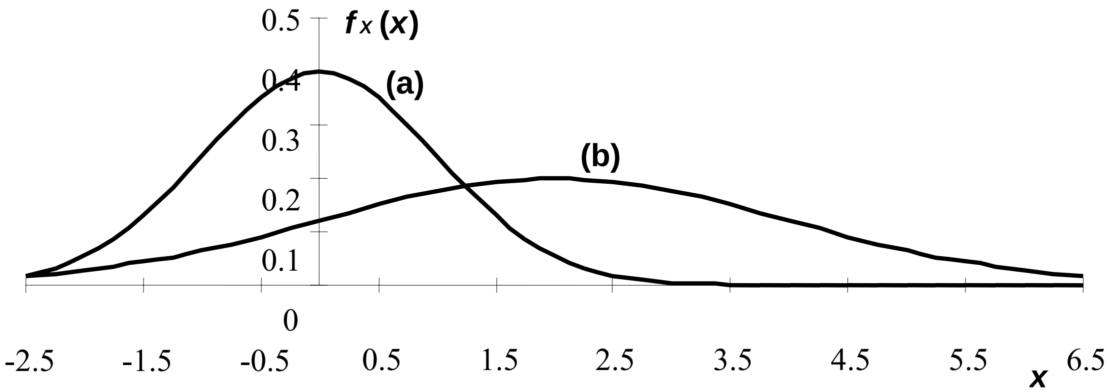

Fig. 2.5 Two examples of normal probability density function (a) $N(0,1)$ and (b) $N(2, 2)$.

2.10.3 A numerical example of the application of the normal distribution

We assume that in an area with wet climate the annual rainfall depth is normally distributed with $\mu = 1750$ mm and $\sigma = 410$ mm. To find the exceedence probability of the value 2500 mm we proceed with the following steps, using the traditional procedure with tabulated z values: $z = (2500 - 1750) / 410 = 1.83$. From normal distribution tables, $F_Z(z) = 0.9664$ ($= F_X(x)$). Hence, $F_X^*(x) = 1 - 0.9664 = 0.0336$.

* For instance, in Excel, the $x \to F_X(x)$ and $F_X(x) \to x$ calculations are done through the functions NormDist and NormInv, respectively (the functions NormSDist and NormSInv can be used for the calculations $z \to F_Z(z)$ and $F_Z(z) \to z$, respectively).

To find the rainfall value that corresponds to exceedence probability 2%, we we proceed with the following steps: $F_X(x) = F_Z(z) = 1 - 0.02 = 0.98$; from the table, $z = 2.05$ hence $x = 1750 + 410 \times 2.05 = 2590.5$ mm. The calculations are straightforward.

2.10.4 The χ^2 distribution

The chi-squared density with n degrees of freedom (symbolically $\chi^2(n)$) is

$$f_X(x) = \frac{1}{2^{n/2}\Gamma(n/2)} x^{n/2-1} e^{-x/2}, \; x \geq 0, \; n = 1, 2, \ldots \tag{2.86}$$

where $\Gamma(\,)$ is the gamma function (not to be confused with the gamma distribution function whose special case is the χ^2 distribution), defined from

$$\Gamma(a) = \int_0^\infty y^{a-1} e^{-y} dy \tag{2.87}$$

The gamma function has some interesting properties such as

$$\begin{array}{lll} \Gamma(1) = 1 & \Gamma(1/2) = \sqrt{\pi} & \Gamma(a+1) = a\Gamma(a) \\ \Gamma(n+1) = n! & \Gamma(n+\frac{1}{2}) = (n-\frac{1}{2})(n-\frac{3}{2})\cdots\frac{3}{2}\sqrt{\pi} & n = 1, 2, \ldots \end{array} \tag{2.88}$$

The χ^2 distribution is a positively skewed distribution (Fig. 2.7) with a single parameter (n). Its mean and variance are n and $2n$, respectively. The coefficients of skewness and kurtosis are $C_s = 2\sqrt{2/n}$ and $C_k = 3 + 12/n$, respectively.

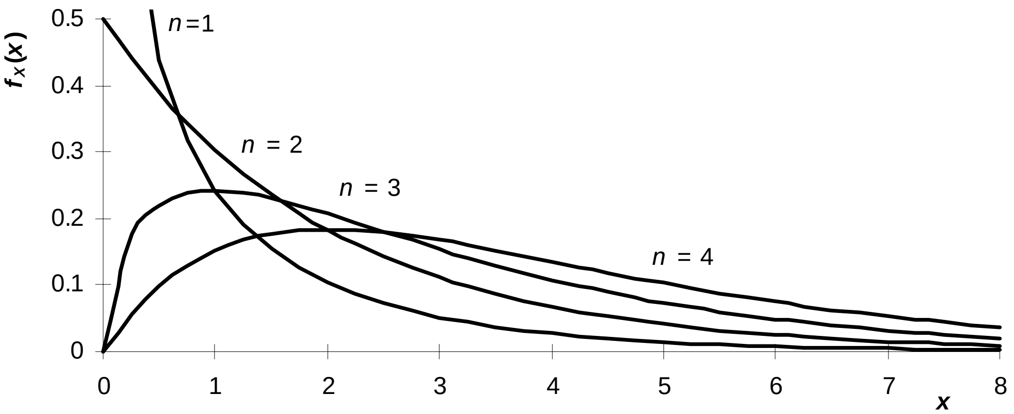

Fig. 2.6 Examples of the $\chi^2(n)$ density for several values of n.

The integral in (2.86) is not calculated analytically, so the typical calculations are based either on tabulated values (see Appendix) or on numerical packages.*

The χ^2 distribution is not directly used to represent hydrological variables; instead the more general gamma distribution (see chapter 6) is used. However, the χ^2 distribution has great importance in statistics (see chapter 3), because of the following theorem: If the random variables X_i ($i = 1, \ldots, n$) are distributed as $N(0, 1)$, then the sum of their squares,

* E.g. in Excel, the relative functions are ChiDist and ChiInv.

$$Q = \sum_{i=1}^{n} X_i^2 \tag{2.89}$$

is distributed as $\chi^2(n)$. Combining this theorem with the central limit theorem we find that for large n the $\chi^2(n)$ distribution tends to the normal distribution.

2.10.5 The Student (*t*) distribution

We shall say that the random variable X has a Student (or t) distribution with n degrees of freedom (symbolically $t(n)$) if its density is

$$f_X(x) = \frac{\Gamma[(n+1)/2]}{\sqrt{\pi n}\,\Gamma(n/2)} \frac{1}{\sqrt{(1 + x^2/n)^{n+1}}}, \quad n = 1, 2, \ldots \tag{2.90}$$

This is a symmetric distribution (Fig. 2.7) with a single parameter (n), mean zero and variance $n/(n-2)$. In contrast to the normal distribution, it has an over-exponential tail but for large n (≥ 30) practically coincides with the normal distribution.

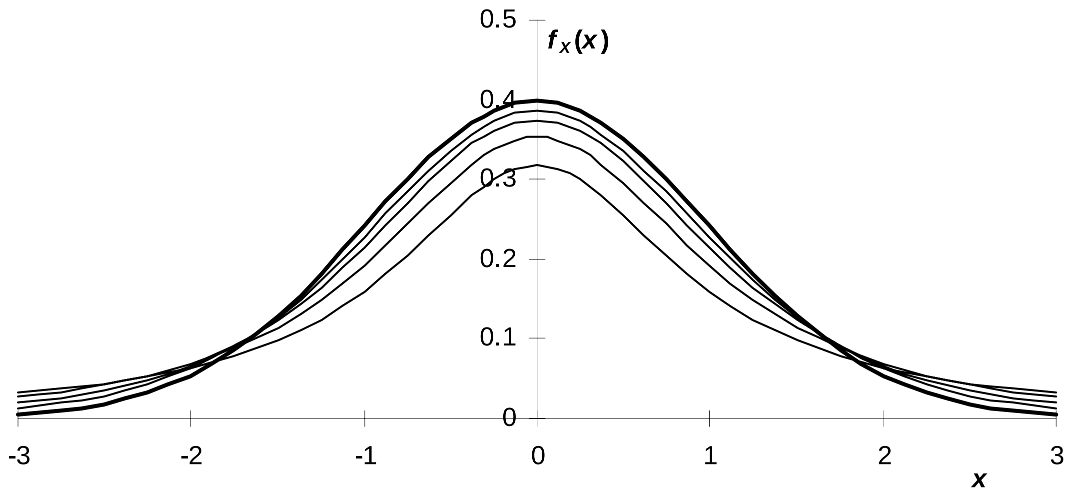

Fig. 2.7 Examples of the $t(n)$ probability density function for n = 1, 2, 4 and 8 (continuous thin lines from down to up), in comparison to the standard normal density $N(0, 1)$ (thick line).

The integral in (2.90) is not calculated analytically, so the typical calculations are based either on tabulated values (see Appendix) or on numerical packages.*

The t distribution is not directly used to represent hydrological variables but it has great importance in statistics (see chapter 3), because of the following theorem: If the random variables Z and W are independent and have $N(0, 1)$ and $\chi^2(n)$ distributions, respectively, then the ratio

$$T = \frac{Z}{\sqrt{W/n}} \tag{2.91}$$

has $t(n)$ distribution.

* E.g. in Excel, the relative functions are TDist and TInv.

References

Greenwood, J. A., J. M. Landwehr, N. C. Matalas, and J. R. Wallis, Probability-weighted moments: Definition and relation to parameters of several distributions expressable in inverse form, *Water Resources Research*, 15, 1049-1054, 1979.

Hosking, J. R. M., L-moments: Analysis and Estimation of Distribution using Linear Combinations of Order Statistics, *Journal of the Royal Statistical Society*, Series B, 52, 105-124, 1990.

Jaynes, E.T., *Probability Theory: The Logic of Science*, Cambridge University Press, 2003.

Kolmogorov, A. N., Foundations of the Theory of Probability, 2nd English Edition, 84 pp. Chelsea Publishing Company, New York, 1956.

Koutsoyiannis, D., Uncertainty, entropy, scaling and hydrological stochastics, 1, Marginal distributional properties of hydrological processes and state scaling, *Hydrological Sciences Journal*, 50(3), 381-404, 2005a.

Koutsoyiannis, D., Uncertainty, entropy, scaling and hydrological stochastics, 2, Time dependence of hydrological processes and time scaling, *Hydrological Sciences Journal*, 50(3), 405-426, 2005b.

Papoulis, A., *Probability and Statistics*, Prentice-Hall, New Jersey, 1990.

Papoulis, A., *Probability, Random Variables, and Stochastic Processes*, 3rd ed., McGraw-Hill, New York, 1991.