CHAPTER 1

Preliminaries

1.1 RANDOM FUNCTIONS

G. de Marsily started the defense of his hydrogeology thesis by showing the audience a jar filled with fine sand and announced "here is a porous medium." Then he shook the jar and announced "and here is another," shook it again and said "and yet another." Indeed, at the microscopic scale the geometry is defined by the arrangement of thousands of individual grains with different shapes and dimensions, and it changes as the grains settle differently each time. Yet at the macroscopic scale we tend to regard it as the same porous medium because its physical properties do not change. This is an ingenious illustration of the notion of a random function in three-dimensional space.

Random functions are useful models for regionalized variables.

1.1.1 Definitions

Notations

Throughout this book the condensed notation x is used to denote a *point* in the n-dimensional space considered. For example, in 3D x stands for the coordinates (x_1, x_2, x_3) (usually called x, y, z). The notation $f(x)$ represents a function of x as well as its value at x. The notation f is used for short, and sometimes the notation $f(\cdot)$ is employed to emphasize that we consider the function taken as a whole and not its value at a single point. Since x is a point in $\mathbb{R}^n$, dx stands for an element of length ($n=1$), of surface ($n=2$), or volume ($n=3$) and $\int_V f(x)dx$ represents the integral of $f(x)$ over a domain $V \subset \mathbb{R}^n$. For example, if $n=2$ and V is the rectangle $[a_1, b_1] \times [a_2, b_2]$, we obtain

$$\int_V f(x)dx = \int_{a_1}^{b_1} dx_1 \int_{a_2}^{b_2} f(x_1, x_2)dx_2$$

Geostatistics: Modeling Spatial Uncertainty, Second Edition. J.P. Chilès and P. Delfiner.

We will seldom need an explicit notation for the coordinates of a point; thus from now on, except when stated otherwise, $x_1, x_2, \ldots$, will represent distinct points in $\mathbb{R}^n$ rather than the coordinates of a single point.

Coming back to the sand jar, we can describe the porous medium by the indicator function of the grains, namely the function $I(x) = 1$ if the point x (in 3D space) is in a grain and $I(x) = 0$ if x is in a void (the pores). Each experiment (shaking the jar) determines at once a whole function $\{I(x) : x \in V\}$ as opposed to, say, throwing a die that only determines a single value (random variable). In probability theory it is customary to denote the outcome of an experiment by the letter ω and the set of all elementary outcomes, or events, by Ω. To make the dependence on the experiment explicit, a random variable is denoted by $X(\omega)$, and likewise our random indicator function is $I(x, \omega)$. For a fixed $\omega = \omega_0$, $I(x, \omega_0)$ is an ordinary function of x, called a *realization* (or *sample function*); any particular outcome of the jar-shaking experiment is a realization of the random function $I(x, \omega)$. On the other hand, for a fixed point $x = x_0$ the function $I(x_0, \omega)$ is an ordinary random variable. Thus mathematically a random function can be regarded as an infinite family of random variables indexed by x.

We can now give a formal definition of a random function [from Neveu (1970), some details omitted; see also Appendix, Section A.1]:

Random Function

Given a domain $D \subset \mathbb{R}^n$ (with a positive volume) and a probability space $(\Omega, \mathcal{A}, P)$, a random function (abbreviation: RF) is a function of two variables $Z(x, \omega)$ such that for each $x \in D$ the section $Z(x, \cdot)$ is a random variable on $(\Omega, \mathcal{A}, P)$. Each of the functions $Z(\cdot, \omega)$ defined on D as the section of the RF at $\omega \in \Omega$ is a realization of the RF. For short the RF is simply denoted by $Z(x)$, and a realization is represented by the lowercase $z(x)$.

In the literature a random function is also called a *stochastic process* when x varies in a 1D space, and can be interpreted as time, and it is called a *random field* when x varies in a space of more than one dimension.

In geostatistics we act *as though* the regionalized variable under study $z(x)$ is a realization of a parent random function $Z(x)$. Most of the time we will not be able to maintain the notational distinction between $Z(x)$ and $z(x)$, and we will get away with it by saying that the context should tell what is meant. The same is true for the distinction between an estimator (random) and an estimate (fixed).

Spatial Distribution

A random function is described by its *finite-dimensional distributions*, namely the set of all multidimensional distributions of k-tuples $(Z(x_1), Z(x_2), \ldots, Z(x_k))$ for all *finite* values of k and all configurations of the points $x_1, x_2, \ldots, x_k$. For short we will call this the *spatial distribution*.

In theory, the spatial distribution is not sufficient to calculate the probability of events involving an infinite noncountable number of points, such as the following important probabilities:

$\Pr\{\sup[Z(x) : x \in V] < z_0\}$ the maximum value in V is less than z_0
$\Pr\{\exists\, x \in V : Z(x) = 0\}$ a zero crossing occurs in domain V
$\Pr\{$every realization of $Z(\cdot)$ is continuous over $V\}$

This difficulty is overcome by adding the assumption of separability of the random function. A random function is separable if all probabilities involving a noncountable number of points can be uniquely determined from probabilities on countable sets of points (e.g., all points in $\mathbb{R}^n$ with rational coordinates), and hence from the spatial distribution. A fundamental result established by Doob (1953, Section 2.2) states that for any random function there always exists a separable random function with the same spatial distribution. In other words, among random functions that are indistinguishable from the point of view of their spatial distribution, we pick and work with the smoothest possible version (see footnote 3 in Section 2.3.1). For completeness let us also mention that tools more powerful than the spatial distribution are required to represent random sets [e.g., Matheron (1975a)] but will not be needed in this book.

Moments

The mean of the RF is the expected value $m(x) = \mathrm{E}[Z(x)]$ of the random variable $Z(x)$ at the point x. It is also called the *drift* of Z, especially when $m(x)$ varies with location. The (centered) covariance $\sigma(x, y)$ is the covariance of the random variables $Z(x)$ and $Z(y)$:

$$\sigma(x,\ y) = \mathrm{E}[Z(x) - m(x)][Z(y) - m(y)]$$

In general, this function depends on both x and y. When $x = y$, $\sigma(x,\ x) = \mathrm{Var}\ Z(x)$ is the variance of $Z(x)$. Higher-order moments can be defined similarly.

Naturally, in theory, these moments may not exist. As usual in probability theory, the mean is defined only if $\mathrm{E}\,|Z(x)| < \infty$. If $\mathrm{E}\,[Z(x)]^2$ is finite at every point, $Z(x)$ is said to be a second-order random function: It has a finite variance, and the covariance exists everywhere.

Convergence in the Mean Square

A sequence of random variables X_n is said to converge in the mean square (m.s.) sense to a random variable X if

$$\lim_{n\to\infty} \mathrm{E}|X_n - X|^2 = 0$$

Taking $X_n = Z(x_n)$ and $X = Z(x)$, we say that an RF $Z(x)$ on $\mathbb{R}^n$ is m.s. continuous if $x_n \to x$ in $\mathbb{R}^n$ implies that $Z(x_n) \to Z(x)$ in the mean square. This definition generalizes the continuity of ordinary functions.

1.1.2 Hilbert Space of Random Variables

It is interesting to cast the study of random functions in the geometric framework of Hilbert spaces. To this end, consider for maximum generality a family

of complex-valued random variables X defined on a probability space $(\Omega, \mathcal{A}, P)$ and having finite second-order moments

$$\mathrm{E}|X|^2 = \int |X(\omega)|^2 P(d\omega) < \infty$$

These random variables constitute a vector space denoted $\mathrm{L}^2(\Omega, \mathcal{A}, P)$ which can be equipped with the scalar product $\langle X, Y\rangle = \mathrm{E}[X\overline{Y}]$ defining a norm[1] (or distance) $\|X\| = \sqrt{\mathrm{E}|X|^2}$ (the upper bar denotes complex conjugation). In this sense we can say that two random variables are *orthogonal* when they are *uncorrelated*. Then $\mathrm{L}^2(\Omega, \mathcal{A}, P)$ is a Hilbert space (every Cauchy sequence converges for the norm). An example is the infinite-dimensional Hilbert space of random variables $\{Z(x) : x \in D\}$ defined by the RF Z.

A fundamental property of a Hilbert space is the possibility of defining the orthogonal projection of X onto a closed linear subspace K as the unique point X_0 in the subspace nearest to X. This is expressed by the so-called *projection theorem* [e.g., Halmos (1951)]:

$$X_0 = \arg\min_{Y \in K} \|X - Y\| \qquad \Leftrightarrow \qquad \langle X - X_0, Y\rangle = 0 \quad \text{for all } Y \in K \tag{1.1}$$

Since $X_0 \in K$, it satisfies $\langle X - X_0, X_0\rangle = 0$ so that

$$\|X - X_0\|^2 = \|X\|^2 - \|X_0\|^2 \tag{1.2}$$

This approximation property is the mathematical basis of kriging theory.

1.1.3 Conditional Expectation

Consider a pair of random variables (X, Y), and let $f(y \mid x)$ be the density of the conditional distribution of Y given that $X = x$. The *conditional expectation of Y given $X = x$* is the mean of that conditional distribution

$$\mathrm{E}(Y \mid X = x) = \int_{-\infty}^{+\infty} y\, f(y \mid x)\, dy$$

$\mathrm{E}(Y \mid X = x) = \phi(x)$ is a function of x only, even though Y appears in the expression. It is also known as the *regression function* of Y on X. When (X, Y) are jointly Gaussian,[2] this function is a straight line. If the argument of $\phi(\cdot)$ is the random variable X, $\phi(X)$ is itself a random variable denoted by $\mathrm{E}(Y \mid X)$. This definition carries over to the case where there are several conditioning variables $X_1, \ldots, X_N$.

[1] Strictly speaking, $\|X\| = 0$ implies that $X = 0$ only up to a set of probability zero, but as usual, equivalence classes of random variables are considered.

[2] "Gaussian" and "normal" will be used as synonyms.

It is possible to develop a theory of conditional expectations without reference to conditional distributions, and this is mathematically better and provides more insight. The idea is to find the best approximation of Y by a function of X. Specifically, we assume X and Y to have finite means and variances and pose the following problem: Find a function $\phi(X)$ such that $\mathrm{E}[Y-\phi(X)]^2$ is a minimum. The solution is the conditional expectation $\mathrm{E}(Y|X)$.

This solution is unique (up to an equivalence between random variables) and is *characterized* by the following property:

$$\mathrm{E}\{[Y-\mathrm{E}(Y\mid X)]H(X)\}=0 \qquad \text{for all measurable } H(\cdot) \tag{1.3}$$

In words, the error $Y-\mathrm{E}(Y\mid X)$ is uncorrelated[3] with any finite-variance random variable of the form $H(X)$. Notice that this is a particular application of the projection formula (1.1).

In particular, when $H(X)\equiv 1$, we get

$$\mathrm{E}[\mathrm{E}(Y\mid X)]=\mathrm{E}(Y) \tag{1.4}$$

The conditional variance is defined by

$$\mathrm{Var}(Y\mid X)=\mathrm{E}(Y^2\mid X)-[\mathrm{E}(Y\mid X)]^2$$

from which we deduce the well-known *total variance* formula

$$\mathrm{Var}(Y)=\mathrm{Var}[\mathrm{E}(Y\mid X)]+\mathrm{E}[\mathrm{Var}(Y\mid X)] \tag{1.5}$$

The variance about the mean equals the variance due to regression plus the mean variance about regression.

For $H(X)=\mathrm{E}(Y\mid X)$ we have

$$\mathrm{E}\{[Y-\mathrm{E}(Y\mid X)]\mathrm{E}(Y\mid X)\}=0$$

so that

$$\mathrm{Cov}(Y,\mathrm{E}(Y\mid X))=\mathrm{Var}(\mathrm{E}(Y\mid X))$$

which shows that Y and $\mathrm{E}(Y\mid X)$ are always positively correlated with

$$\rho^2=\frac{\mathrm{Var}(\mathrm{E}(Y\mid X))}{\mathrm{Var}(Y)} \tag{1.6}$$

From (1.5) the residual variance takes the familiar form

[3] This does not imply independence between the error and X; if $X=Y^2$, Y symmetric about 0, $\mathrm{E}(Y\mid X)=0$, but Y is not independent of X.

$$\mathrm{E}[\mathrm{Var}(Y \mid X)] = (1 - \rho^2)\mathrm{Var}(Y) \tag{1.7}$$

(note that here ρ is not the correlation between Y and X but between Y and its regression on X).

In addition to the unbiasedness property, let us mention the property of *conditional unbiasedness*, which we will often invoke in this book in relation to kriging:

$$\phi(X) = \mathrm{E}(Y \mid X) \qquad \Rightarrow \qquad \mathrm{E}(Y \mid \phi(X)) = \phi(X)$$

The proof follows immediately from the characteristic property (1.3), since

$$\mathrm{E}\{[Y - \phi(X)]H(X)\} = 0 \qquad \text{for all measurable } H(\cdot)$$

entails that $\phi(X)$ also satisfies

$$\mathrm{E}\{[Y - \phi(X)]H(\phi(X))\} = 0 \qquad \text{for all measurable } H(\cdot)$$

Some Properties of Conditional Expectation

The following results can be derived directly from the characteristic formula and are valid almost surely (a.s.):

Linearity	$\mathrm{E}(aY_1 + bY_2 \mid X) = a\,\mathrm{E}(Y_1 \mid X) + b\,\mathrm{E}(Y_2 \mid X)$
Positivity	$Y \geq 0$ a.s. $\Rightarrow$ $\mathrm{E}(Y \mid X) \geq 0$ a.s.
Independence	X and Y are independent $\Rightarrow$ $\mathrm{E}(Y \mid X) = \mathrm{E}(Y)$
Invariance	$\mathrm{E}(Y f(X) \mid X) = f(X)\ \mathrm{E}(Y \mid X)$
Successive projections	$\mathrm{E}(Y \mid X_1) = \mathrm{E}[\mathrm{E}(Y \mid X_1, X_2) \mid X_1]$

1.1.4 Stationary Random Functions

Strict Stationarity

A particular case of great practical importance is when the finite-dimensional distributions are invariant under an arbitrary translation of the points by a vector h:

$$\Pr\{Z(x_1) < z_1, \ldots, Z(x_k) < z_k\} = \Pr\{Z(x_1 + h) < z_1, \ldots, Z(x_k + h) < z_k\}$$

Such RF is called *stationary*. Physically, this means that the phenomenon is homogeneous in space and, so to speak, repeats itself in the whole space. The sand in the jar is a good image of a stationary random function in three dimensions, at least if the sand is well sorted (otherwise, if the jar vibrates, the finer grains will eventually seep to the bottom, creating nonstationarity in the vertical dimension).

Second-Order Stationarity

When the random function is stationary, its moments, if they exist, are obviously invariant under translations. If we consider the first two moments only, we have for points x and $x+h$ of $\mathbb{R}^n$

$$\begin{aligned} &\mathrm{E}[Z(x)] = m, \\ &\mathrm{E}[Z(x) - m][Z(x+h) - m] = C(h) \end{aligned}$$

The mean is constant and the covariance function only depends on the *separation* h. We will see in Section 2.3.2 that a covariance must be a *positive definite* function.

By definition, a random function satisfying the above conditions is *second-order stationary* (or weakly stationary, or wide-sense stationary). In this book, unless specified otherwise, stationarity will always be considered at order 2, and the abbreviation SRF will designate a second-order stationary random function.

An SRF is *isotropic* if its covariance function only depends on the length $|h|$ of the vector h and not on its orientation.

Intrinsic Hypothesis

A milder hypothesis is to assume that for every vector h the *increment* $Y_h(x) = Z(x+h) - Z(x)$ is an SRF in x. Then $Z(x)$ is called an *intrinsic random function* (abbreviation: IRF) and is characterized by the following relationships:

$$\begin{aligned} &\mathrm{E}[Z(x+h) - Z(x)] = \langle a, h \rangle, \\ &\mathrm{Var}[Z(x+h) - Z(x)] = 2\gamma(h) \end{aligned}$$

$\langle a, h \rangle$ is the *linear drift* of the IRF (drift of the increment) and $\gamma(h)$ is its *variogram* function, studied at length in Chapter 2.

If the linear drift is zero—that is, if the mean is constant—we have the usual form of the intrinsic model:

$$\begin{aligned} &\mathrm{E}[Z(x+h) - Z(x)] = 0, \\ &\mathrm{E}[Z(x+h) - Z(x)]^2 = 2\gamma(h) \end{aligned}$$

Gaussian Random Functions

A random function is Gaussian if all its finite-dimensional distributions are multivariate Gaussian. Since a Gaussian distribution is completely defined by its first two moments, knowledge of the mean and the covariance function suffices to determine the spatial distribution of a Gaussian RF. In particular, second-order stationarity is equivalent to full stationarity.

A Gaussian IRF is an IRF whose increments are multivariate Gaussian.

A weaker form of Gaussian behavior is when all *bivariate* distributions of the RF are Gaussian; the RF is then sometimes called *bi-Gaussian*. A yet weaker form

is when only the marginal distribution of $Z(x)$ is Gaussian. This by no way implies that $Z(x)$ is a Gaussian RF, but this leap of faith is sometimes made.

1.1.5 Spectral Representation

The spectral representation of SRFs plays a key role in the analysis of time signals. It states that a stationary signal is a mixture of statistically independent sinusoidal components at different frequencies. These basic harmonic constituents can be identified physically by means of filters that pass oscillations in a given frequency interval and stop others. This can also be done digitally using the discrete Fourier transform.

In the case of spatial processes the physical meaning of frequency components is generally less clear, but the spectral representation remains a useful theoretical tool, especially for simulations. For generality and in view of future reference we will state the main results in $\mathbb{R}^n$, which entails some unavoidable mathematical complication.

Theorem. A real, continuous, zero-mean RF defined on $\mathbb{R}^n$ is stationary (of order 2) if and only if it has the spectral representation

$$Z(x) = \int e^{2\pi i\langle u,x\rangle}\, Y(du) \tag{1.8}$$

for some unique *orthogonal random spectral measure* $Y(du)$ (see Appendix, Section A.1). Here i is the unit pure imaginary number, $u=(u_1,\ldots,u_n)$ denotes an n-dimensional frequency vector, du is an element of volume in $\mathbb{R}^n$, $x=(x_1,\ldots,x_n)$ is a point of $\mathbb{R}^n$, and $\langle u,\ x\rangle = u_1x_1+\cdots+u_n x_n$ is the scalar product of x and u.

For any Borel sets B and B' of $\mathbb{R}^n$, the measure Y satisfies

$$\begin{aligned} &\mathrm{E}[Y(B)] = 0 \\ &\mathrm{E}[Y(B)\overline{Y(B')}] = 0 \qquad\qquad \text{if} \quad B\cap B' = \varnothing \\ &Y(B\cup B') = Y(B) + Y(B') \quad \text{if} \quad B\cap B' = \varnothing \end{aligned}$$

$Z(x)$ being real, we have in addition the symmetry relation $Y(-B) = \overline{Y(B)}$, where $-B$ denotes the symmetric of B with respect to the origin. Note that the random variables associated with disjoint sets B and B' are uncorrelated, hence the name *orthogonal* measure.

Now define $F(B) = \mathrm{E}|Y(B)|^2$. F is a positive bounded symmetric measure called the *spectral measure*. We have in particular

$$\begin{aligned} &F(B\cup B') = F(B) + F(B') \quad \text{if} \quad B\cap B' = \varnothing \\ &\mathrm{E}[Y(B)\overline{Y(B')}] = F(B\cap B') \end{aligned}$$

It follows readily from (1.8) and the symmetry of F that the covariance of $Z(x)$ has the spectral representation

$$C(h) = \mathrm{E}[Z(x)\overline{Z(x+h)}] = \int e^{2\pi i\langle u,h\rangle} F(du)$$

For time signals, the power of the RF $Z(x)$, which is the energy dissipated per unit time, is generally proportional to $Z(x)^2$. If the SRF has zero mean, $C(0)$ is equal to $\mathrm{E}[Z(x)^2]$ and plays the role of an average power, and the measure F represents the decomposition of this power into the different frequencies. Note that the integral $\int F(du)$ of the spectral measure is equal to the total power $C(0)$.

Real Spectral Representation

It is interesting to separate the real and imaginary parts of the random spectral measure Y in the form

$$Y(B) = U(B) - i\ V(B)$$

where U and V are two *real* random measures (notice the $-i$ in the definition of V). From the properties of Y, we can deduce the following properties that will be useful for simulations:

$$\begin{array}{ll} U(-B) = U(B) & V(-B) = -V(B) \\ \mathrm{E}[U(B)U(B')] = \mathrm{E}[V(B)V(B')] = 0 & \text{if} \quad B \cap B' = B \cap (-B') = \varnothing \\ \mathrm{E}[U(B)V(B')] = 0 & \forall B, B' \\ \mathrm{E}\left[U(B)^2\right] = \mathrm{E}\left[V(B)^2\right] = F(B)/2 & \text{if} \quad \{0\} \notin B \\ \mathrm{E}\left[|U(\{0\})|^2\right] = F(\{0\}) & V(\{0\}) = 0 \end{array} \tag{1.9}$$

Also, $Z(x)$ has the representation

$$Z(x) = \int \cos(2\pi\langle u, x\rangle) U(du) + \int \sin(2\pi\langle u, x\rangle) V(du)$$

1.1.6 Ergodicity

Ergodicity is an intimidating concept. The practitioner has heard that the RF should be ergodic, since "this is what makes statistical inference possible," but he or she is not sure how to check this fact and proceeds anyway, feeling vaguely guilty of having perhaps overlooked something very important. We will attempt here to clarify the issues. In practice, ergodicity is never a problem. When no replication is possible, as with purely spatial phenomena, we can safely choose an ergodic model. If the phenomenon is repeatable, typically time-dependent fields or simulations, averages are computed over the different realizations, and the only issue (more a physical than a mathematical one) is to make sure that we are not mixing essentially different functions.

A detailed discussion of ergodicity can be found in Yaglom (1987, Vol. I, Chapter 3), and an analysis of its meaning in the context of unique phenomena in Matheron (1978). We have summarized the most important results so that practitioners can pay their respects to ergodicity once for all and move on.

Ergodic Property

In order to carefully distinguish a random function from its realizations, we will revert, in this section only, to the full notation $Z(x, \omega)$, where ω is the random event indexing the realization. By definition, a stationary random function $Z(x, \omega)$ is ergodic (in the mean) if the spatial average of $Z(x, \omega)$ over a domain $V \subset \mathbb{R}^n$ converges to the expected value $m = \mathrm{E}[Z(x, \omega)]$ when V tends to infinity:

$$\lim_{V \to \infty} \frac{1}{|V|} \int_V Z(x, \omega) dx = m \tag{1.10}$$

In this expression the norming factor $|V|$ denotes the volume of the domain V, and the limit is understood, as we will always do, in the mean square sense. In $\mathbb{R}^n$ it is important to specify how V tends to infinity, since we may imagine that V becomes infinitely long in some directions only, but we exclude this and assume that V grows in *all* directions. For example, V may be the cube $[0, t]^n$, where $t \to \infty$. Of course the limit does not depend on the particular shape of V.

To gain insight into the meaning of this property, it is interesting to revisit the sand jar a last time and do a little thought experiment. We consider a point x at a fixed location relative to the jar and shake the jar repeatedly, recording each time a 1 if x falls in a grain and a 0 otherwise. From this we can evaluate the mean of $I(x, \omega)$, namely the probability that x is in a grain, which should not depend on x. It is intuitively obvious that we will get the same result if we keep the jar fixed and select the point x at random within the jar, the probability of landing in a grain being equal to the proportion of the space occupied by the grains.

The ergodic property can be extremely important for applications, since it allows the determination of the mean from a *single* realization of the stationary random function, and precisely most of the time we only have one realization to work with. Not all stationary random functions are ergodic. The classic counterexample is the RF $Z(x, \omega) \equiv A(\omega)$ whose realizations are constants drawn from the random variable A. Clearly for each realization the space integral (1.10) is equal to the constant level $A(\omega)$ but not to the mean of A. Another, more realistic example of nonergodic RF is to consider a family of different stationary and ergodic RFs and select one of them according to the outcome of some random variable A, thus defining the composite RF $Z(x, \omega; A)$. On each realization the space integral converges to the mean $m(a) = \mathrm{E}(Z(x, \omega; A) \mid A = a)$ of the particular RF $Z(x, \omega; a)$, but this is different from the overall mean $\mathrm{E}[m(A)]$. Here we have the most common source of stationary but nonergodic random functions arising in practice. As has been pointed out in the literature, nonergodicity usually means that the random function comprises an artificial union of a number of distinct ergodic stationary functions.

Ergodic Theorem

This theorem states that if $Z(x, \omega)$ is a stationary random function (of order 2), the space integral (1.10) *always* converges to some value $m(\omega)$, but this value in general depends on the realization ω: it is a random variable, not a constant:

$$\lim_{V \to \infty} \frac{1}{|V|} \int_V Z(x, \omega) dx = m(\omega) \tag{1.11}$$

This result is a direct consequence of the stationarity of $Z(x, \omega)$ and again requires V to grow in all directions. The random variable $m(\omega)$ has mean m and a fluctuation equal to the atom at the origin of the random spectral measure associated with the RF $Z(x, \omega)$:

$$m(\omega) = m + Y(\{0\}, \omega)$$

Since $\mathrm{E}|Y(\{0\},\omega)|^2 = F(\{0\})$, it appears that $Z(x,\ \omega)$ possesses the ergodic property if and only if its spectral measure F has no atom at the origin.

An equivalent condition, known as Slutsky's ergodic theorem, is

$$\lim_{V\to\infty}\frac{1}{|V|}\int_V C(h)dh = 0 \tag{1.12}$$

This is always satisfied if $C(h) \to 0$ as $h \to \infty$, as is usually the case, but the condition is not necessary. Moreover, if the integral of the covariance in $\mathbb{R}^n$ is finite, when V is very large the left-hand side of (1.12) is an approximation to the variance of the space integral (1.11), and we will revisit this in Section 2.3.5 with the notion of *integral range*.

Ergodicity in the Covariance

In the above we have only considered first-order ergodicity, or ergodicity in the mean. It is also important to be able to determine the covariance from a single realization. This implies second-order ergodicity, or ergodicity in the covariance. To establish the ergodicity of the covariance, the same theory can be applied to the product variable $Q_h(x) = Z(x)\ Z(x+h)$ considered as a second-order stationary random function of x with h fixed. This involves the stationarity of fourth-order moments. For Gaussian RFs the fourth-order moments depend on the second-order moments, and simple results can be obtained. The analogue of Slutsky's condition for the convergence of covariance estimates is then

$$\lim_{V\to\infty}\frac{1}{|V|}\int_V [C(h)]^2 dh = 0 \tag{1.13}$$

This condition is more restrictive than (1.12). Its equivalent spectral formulation is that the spectral measure F has no atom *anywhere*. In other words, the covariance has no sinusoidal component. Again the convergence $C(h) \to 0$ as $h \to \infty$ suffices to fulfill (1.13), but the proof is only valid for Gaussian RFs.

Now What?

In the case of a unique phenomenon, there is no way of knowing if the space integral would have converged to a different value on another realization, since there is, and can be, only one. As will be seen in a moment, ergodicity is not an objective property in the sense that it cannot be falsified. Therefore we *choose* to model $Z(x,\ \omega)$ as an ergodic random function whose mean is the limit of the space integral (1.10). Likewise, we take the limit of the regional covariance (a space integral) as the *definition* of the covariance of the parent RF. Any other choice would have no relevance to the situation considered.

Strictly speaking, there still is a problem. Recall that in practice, we work in a bounded domain and cannot let it tend to infinity. This is a matter of scale. If the domain is large enough for the integral (1.12) to be small, the mean can be

estimated reliably. But if the variance is still large, due to a slow fall-off of the covariance, the estimation of the mean is difficult, and it is preferable to avoid using it at all and only consider increments. This is the justification for using the variogram instead of the covariance. The possibility of statistical inference of the variogram is discussed in Section 2.9.

When dealing with space–time phenomena observed at a fixed set of monitoring stations, we typically consider spatially nonstationary models and compute time-averaged estimates of spatial means and covariances. We are thus treating the data as a collection of (correlated) stationary and ergodic random functions of time (multiple time series). These assumptions have to be checked carefully.

Micro-Ergodicity

As we have noted earlier, it is impossible to extend the domain to infinity. Matheron (1978) introduced the notion of *micro-ergodicity*, also called *infill asymptotics* (Cressie, 1991), concerned with the convergence of space integrals when the domain D remains fixed but the sampling density becomes infinite. This concept is distinct from standard ergodicity. For example, neither the mean nor the variance of a stationary and ergodic RF is micro-ergodic, but the slope of the variogram at the origin is micro-ergodic if the variable is not too smooth (Section 2.9.2). Micro-ergodic parameters represent physically meaningful properties.

1.2 ON THE OBJECTIVITY OF PROBABILISTIC STATEMENTS

What sense does it make to speak of the probability of a unique event? When we are told that "there is a 60% chance of rain tomorrow," we know the next day if it rains or not, but how can we check that the probability of rain was indeed 60% on that day at a specific place? We can't. The only probabilistic statement that can be disproved is "there is a zero chance of rain tomorrow": If it does rain the next day, then clearly the forecast was wrong. The same problem essentially arises for spatial "prediction." What is the physical meaning of a statement such as "there is a 0.95 probability that the average porosity of this block is between 20% and 25%" ? Potentially we could measure the porosity and check if it lies in the interval, but we will never know if the 0.95 was correct. Yet despite their unclear meaning, we tend to find probabilistic statements useful in giving us an appreciation of uncertainty.

In reality we establish the credibility of weather forecasts not from a single prediction but over time. Someone with enough motivation could check if out of 100 days associated with a forecast of a 60% chance of rain about 60 days were indeed rainy, and do this for all % chance classes. A successful track record, without proving the correctness of the forecast on any given day, proves at least that it is correct *on the average*. It validates the forecasting *methodology*.

One may object that since we introduced repetitions we are no longer really dealing with a unique phenomenon. But the distinction between unique and

repeatable situations is not as clear-cut as it seems. Strictly speaking, it is impossible to repeat the "same" experiment: They always differ in some aspects; we simply judge those unimportant. On the other hand, even though every petroleum reservoir, every mine, and every forest is unique, they all belong to classes of situations, shaly sand reservoirs, copper deposits, or tropical woods that are similar enough to give rise to specific methodologies that over time can be validated objectively. This is "external" objectivity.

The practitioner who is interested in the evaluation of this specific deposit or that specific forest would rather have criteria for "internal" objectivity that are based on those unique situations. If we cannot pin down the meaning of a probabilistic statement on a singular event, then the question becomes, Which concepts, statements, and parameters have an objective, observable, measurable counterpart in reality? Matheron (1978) devotes a fascinating essay entitled "Estimating and Choosing" to this quest for objectivity. The central idea is this: *The only objective quantities are those that may be calculated from the values of a single realization over a bounded domain D*. Indeed, in the absence of repetitions, the maximum information we can ever get is the complete set of values $\{z(x) : x \in D\}$. Objective quantities are essentially space integrals of functions of $z(x)$, referred to as *regionals*: all the values of $z(x)$ itself, block averages, mean values above thresholds, and so on, along with the regional mean, variogram, or histogram. On the contrary, the expected value m, the (true) variogram γ, or the marginal distribution of $Z(x)$ are *conventional* parameters. To emphasize the difference, Matheron says that we *estimate* a regional whose exact value is unknown but nevertheless exists independently of us, namely is potentially observable, but we *choose* the value of a conventional parameter.[4]

These considerations lead to a striking reversal of point of view where regionals cease to be mere estimates of "true" parameters to become the physical reality itself, while their theoretical counterparts turn into conventional parameters. For example, the regional variogram γ_R should not be regarded as the regional version of γ but rather γ as being the theoretical version of γ_R. Likewise, the fluctuation variance (in the probabilistic model) of a regional is not indicative of the difficulty of the statistical inference of its expected value but rather of the lack of objective meaning of this parameter.

The objectivity of statements can be defined by two criteria. The stronger one is to regard a statement as objective if it is *decidable*, which means that it can be declared true or false once we know $z(x)$ for all $x \in D$. The weaker form of objectivity is K. Popper's demarcation criterion for scientific hypotheses: It must be possible to design experiments whose outcomes are liable to *falsify* predictions derived from these hypotheses, that is, events with probabilities (nearly) equal to 0 or 1 (in the model). If such attempts are successful, the

[4] In the statistical literature, *to predict* means *to estimate* in Matheron's sense and *to estimate* means *to choose*. In this book we will estimate observables in Matheron's sense but *fit* model parameters rather than choose them.

hypothesis is falsified. If it withstands testing, we may not conclude that it is true but only that it is corroborated (not refuted).

The statement "$z(x)$ is a realization of a random function $Z(x)$" or even "of a stationary random function" has no objective meaning. Indeed, since D is bounded, it is always possible by periodic repetitions and randomization of the origin to construct a stationary random function having a realization that coincides over D with the observed $z(x)$. Therefore no statistical test can disprove stationarity in general. We *choose* to consider $z(x)$ as a realization of $Z(x)$ over D. It does not mean that this decision is arbitrary—in practice, it is suggested by the spatial homogeneity of the data—but simply that it cannot be refuted. As stated earlier, ergodicity is also not an objective property.

If repetitions are the objective foundation of probabilities and if only regionals are physically meaningful, then in the case of a unique phenomenon the objectivity of our measures of uncertainty must be based on spatial repetitions. These are obtained by moving the configuration involved—for example, a block and its estimating data points—throughout a domain D_0. Denoting by Z_v the block value and by Z^* its estimator, the estimation variance of such block is interpreted as the spatial average of the squared error $(Z^* - Z_v)^2$ over D_0. By construction, this variance is *not localized* (i.e., is constant) within D_0, but neither is the kriging variance calculated in the stationary model (since it only depends on the geometry and on the variogram). Of course the domain D_0 can itself be local and correspond to a homogeneous subzone of the total domain D. However, D_0 should not be too small; otherwise, we will lack repetitions. We are tempted to say that there is a trade-off between objectivity and spatial resolution.[5]

1.3 TRANSITIVE THEORY

To avoid the epistemological problems associated with the uniqueness of phenomena, Matheron (1965) first developed an estimation theory in purely spatial terms which he named *transitive theory*. In this approach the regionalized variable $z(x)$ is deterministic and only assumed to be identically zero outside a bounded domain D; it represents a so-called transition phenomenon, a spatial equivalent of a transient phenomenon in time. We will focus here on the global estimation problem, namely the evaluation of the integral of $z(x)$ which typically represents the total amount of some resource. Initially applied to mineral resources, the transitive approach has received a renewed interest in the last two decades for the estimation of fish abundance when the areas of fish presence have diffuse limits (Petitgas, 1993; Bez, 2002).

[5] The image of "dithering" comes to mind. This is a binarization technique to transform a halftone image into a black-and-white image. A gray level is obtained by judiciously distributing black and white dots in the cells of a matrix: A 4×4 matrix allows 16 gray levels, and a 16×16 matrix is required to render 256 gray levels. Thus there is a trade-off between the representation of gray level amplitude and the spatial resolution.

In this model, randomness is introduced through sampling. The easiest would be to use the classic Monte Carlo method and select N samples randomly and independently, leading to an unbiased estimator with a variance equal to σ^2/N, where σ^2 is the spatial variance of $z(x)$. However, systematic sampling is usually more efficient. We will present the theory in this case, mainly for background, but also to justify a neat formula for surface estimation. The transitive theory can also be developed for local estimation, but it has no advantage over the more elegant random function approach. Transitive theory will not be used elsewhere in the book.

1.3.1 Global Estimation by Systematic Sampling

Consider the estimation of the integral

$$Q = \int z(x)dx$$

which is finite since $z(x)$ is zero outside the domain D. If $z(x)$ is a mineral grade (in g/ton) and if the ore density d is a constant, Qd is the quantity of metal in the deposit; if $z(x)$ is an indicator function, Q is the volume of D. We assume that the domain D is sampled on a rectangular grid that extends as far as needed beyond the boundaries of D.

As usual, we reason in $\mathbb{R}^n$ and denote by a the elementary grid spacing $(a_1, a_2, \ldots, a_n)$ and by $|a|$ its volume (i.e., the product $a_1\ a_2 \ldots a_n$). The origin of the grid, which is one of its points, is denoted by x_0, and k denotes the set of positive or negative integers $k = (k_1, k_2, \ldots, k_n)$. The simplest estimate of Q is

$$Q^*(x_0) = |a| \sum_{k \in \mathbb{Z}^n} z(x_0 + ka)$$

where $\mathbb{Z}$ is the set of relative integers. If we select the origin x_0 at random and uniformly within the parallelepiped $\Pi = [0,\ a_1] \times [0,\ a_2] \times \cdots \times [0,\ a_n]$, then Q^* becomes a random variable whose expected value is

$$\mathrm{E}(Q^*) = \frac{1}{|a|} \int_\Pi Q^*(x_0)dx_0 = \frac{1}{|a|} \int_\Pi dx_0 |a| \sum_k z(x_0 + ka) = \int z(x)dx = Q$$

It is unbiased. When we define the *transitive covariogram* $g(h)$ by

$$g(h) = \int z(x)z(x+h)\,dx$$

similar calculations show that

$$\mathrm{E}(Q^*)^2 = |a| \sum_k g(ka) \qquad \text{and} \qquad Q^2 = \int g(h)\,dh$$

so that the variance of the error $Q^* - Q$, or *estimation variance*, is given by the formula

$$\mathrm{E}(Q^* - Q)^2 = |a| \sum_k g(ka) - \int g(h)\,dh \tag{1.14}$$

This estimation variance, denoted $\sigma^2(a)$, appears as the error incurred by approximating the integral $\int g(h)\,dh$ by a discrete sum over the grid. It decreases as the grid becomes finer and as the function $z(\cdot)$ becomes smoother.

This variance is always nonnegative provided that the covariogram $g(h)$ is modeled as a positive definite function ($g(h)$ is the convolution of $z(x)$ by $z(-x)$). The transitive covariogram plays the role of the covariance in an RF model, and in fact they are related, since the regional *noncentered* covariance over D is given by

$$C_{\mathrm{R}}(h) = \frac{1}{K(h)} \int_{D \cap D_{-h}} z(x) z(x+h) dx, \qquad \text{where } K(h) = |D \cap D_{-h}|$$

so that $g(h) = K(h)\ C_{\mathrm{R}}(h)$.

An expansion of formula (1.14) as a Euler–MacLaurin series leads, for small a, to a decomposition of the variance $\sigma^2(a)$ into two terms:

$$\sigma^2(a) = T_1(a) + T_2(a)$$

The first term, $T_1(a)$, is related to the behavior of $g(h)$ near the origin; the second one, $T_2(a)$, depends on its behavior near the range (the distance b beyond which $g(h)$ becomes identically zero). $T_2(a)$ is the *fluctuating term*, also called *Zitterbewegung* (the German for jittery motion). It is a periodic function of the remainder ε of the integer division b/a and cannot be evaluated from the grid data; since it has a zero mean, it is simply ignored.

The regular term $T_1(a)$ can be approximated from the expansion of $g(h)$. For example, in 2D and for an isotropic covariogram with a linear behavior near the origin, the explicit result is

$$\sigma^2(a) = \sigma^2(a_1, a_2) \approx -g'(+0) \left[\frac{1}{6} a_1^2 a_2 + 0.0609 a_2^3 \right], \qquad a_1 \leq a_2 \tag{1.15}$$

where $g'(+0)$ is the slope of $g(h)$ at the origin and is < 0.

1.3.2 Estimation of a Surface Area

If $z(x)$ is the indicator function of a geometric object, its covariogram is necessarily linear at the origin. More precisely, for a vector h in a direction α, we have $g(h) = g(0) - D_\alpha |h| + \cdots$, where D_α is the "total diameter" in the direction α. If the object is convex, D_α is simply the so-called tangent diameter (or caliper diameter); otherwise, $2D_\alpha$ is the total length of the contour of the object projected orthogonally along the direction α (see Section 2.3.4). Here we consider an object with surface area A and a total diameter D that is approximately the same in all directions. Replacing A by its estimate $A^* = N\ a_1\ a_2$, where N is the number of positive samples, we can express the variance (1.15) in the dimensionless form σ_A^2 / A^2:

$$\frac{\sigma_A^2}{A^2} \approx \frac{D}{\sqrt{A}} \frac{1}{N^{3/2}} \left[\frac{1}{6} \sqrt{\lambda} + 0.0609 \lambda^{-3/2} \right], \qquad \lambda = a_1/a_2 \leq 1 \tag{1.16}$$

Note that the variance decreases like $1/N^{3/2}$ rather than $1/N$. We can evaluate D from the contour of the object by counting the number of boundary segments $2N_1$ and $2N_2$ respectively, parallel to a_1 and a_2, including possible holes in the contour (the total perimeter comprises $2\,(N_1 + N_2)$ segments):

$$D = N_1 a_1 = N_2 a_2$$

and upon replacement in (1.16), we get

$$\frac{\sigma_A^2}{A^2} \approx \frac{1}{N^2} \left[\frac{1}{6} N_2 + 0.0609 \frac{N_1^2}{N_2} \right], \qquad N_2 \leq N_1 \tag{1.17}$$

This formula remains valid if the object is not isotropic but has a main direction of elongation parallel to one of the grid axes, which is the natural orientation for the grid. Indeed we can

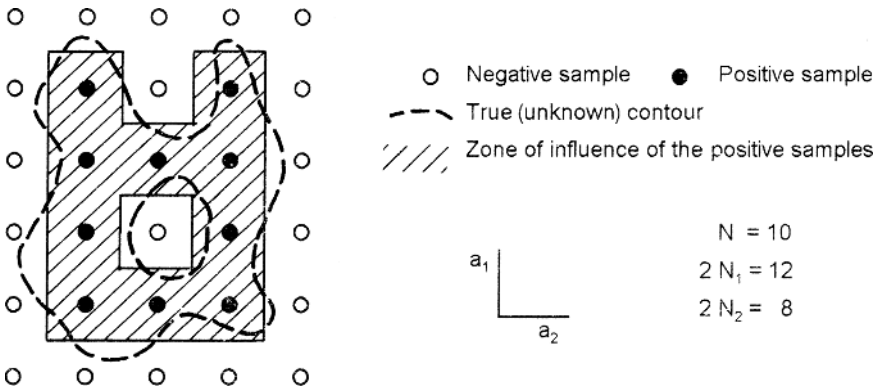

FIGURE 1.1 Estimation of a surface area by systematic sampling.

then restore isotropy, at least approximately, by an affine transformation parallel to one of the grid axes; this changes a_1 or a_2 as well as A and D_1 or D_2 but not N, N_1, N_2, or σ_A^2/A^2, which are dimensionless. Thus (1.17) is a simple, self-contained (no calculation and modeling of $g(h)$ needed), and yet theoretically founded formula for evaluating the error in the estimation of a surface area.

To illustrate its use, consider the example shown in Figure 1.1. We read from the figure:

$$\begin{matrix} N = 10 \\ 2N_1 = 12 \\ 2N_2 = 8 \end{matrix} \quad \text{so that} \quad \frac{\sigma_A^2}{A^2} = \frac{1}{100}\left[\frac{4}{6} + 0.0609\,\frac{36}{4}\right] = \frac{1.21}{100}$$

The relative error standard deviation on the surface area is therefore 11%.

An interesting indication can be derived concerning the optimal grid mesh. In case of an isotropic object the variance in (1.16) is minimized for $\lambda = 1$, that is, $a_1 = a_2$. If the object is not isotropic an affine transformation is applied to restore isotropy, for example multiplying lengths along D_2 by D_1/D_2. The new grid spacings become $a_1' = a_1$ and $a_2' = (D_1/D_2)a_2$, and in this isotropic case the optimal grid mesh satisfies $a_1' = a_2'$. Therefore optimum sampling is achieved when $D_1/a_1 = D_2/a_2$, or equivalently $N_1 = N_2$, that is, when the grid mesh is adapted to the anisotropy of the object. Formula (1.17) also shows that for the optimal grid mesh ($N_1 = N_2$) the estimation variance increases with the perimeter of the object.