3. Spatially-Dependent Random Effects

Observe that all regressions in the illustrations above [starting with expression (2.1.3) in the Sudan rainfall example] have relied on an implicit model of unobserved random effects (i.e., regression residuals) as a collection $(\varepsilon_i : i = 1,..,n)$ of *independently and identically distributed normal random variables* [where for our purposes, individual sample points i are taken to represent different spatial locations, s_i]. But recall from the introductory discussion in Section 1.2 above that for more realistic spatial statistical models we must allow for possible spatial dependencies among these residuals. Hence the main objective of the present section is to extend this model to one that is sufficiently broad to cover the types of spatial dependencies we shall need. To do so, we begin in Section 3.1 by examining random effects at a single location, and show that normality can be motivated by the classical *Central Limit Theorem*. In Section 3.2 , these results will be extended to random effects at multiple locations by applying the Multivariate Central Limit Theorem to motivate multivariate normality of such joint random effects. This *multi-normal model* will form the statistical underpinning for all subsequent analyses. Finally in Section 3.3 we introduce the notion of *spatial stationarity* to model covariances among these spatial random effects $(\varepsilon_i : i = 1,..,n)$.

3.1 Random Effects at a Single Location

First recall that the unobserved random effects, ε_i, at each location (or sample point), s_i, are assumed to fluctuate around zero, with $E(\varepsilon_i) = 0$. Now imagine that this overall random effect, ε_i, is composed of many independent factors,

$$(3.1.1) \qquad \varepsilon_i \;=\; e_{i1} + e_{i2} + \cdots + e_{im} = \sum_{k=1}^{m} e_{ik} \quad ,$$

where in typical realizations some of these factors, e_{ik}, will be positive and others negative. Suppose moreover that each individual factor contributes only a very small part of total. Then no matter how these individual random factors are distributed, their cumulative effect, ε_i, must eventually have a "bell shaped" distribution centered around zero. This can be illustrated by a simple example in which each random component, e_{ik}, assumes the values $1/m$ and $-1/m$ with equal probability, so that $E(e_{ik}) = 0$ for all $k = 1,..,m$. Then each is distributed as shown for the $m = 1$ case in Figure 3.1(a) below. Now even though this distribution is clearly flat, if we consider the $m = 2$ case

$$(3.1.2) \qquad \varepsilon_i \;=\; e_{i1} + e_{i2}$$

then it is seen in Figure 3.1(b) that the distribution is already starting to be "bell shaped" around zero. In particular the value 0 is much more likely than either of the extremes, -1 and 1. The reason of course is that this value can be achieved in *two* ways, namely $(e_{i1} = \frac{1}{2}, e_{i2} = -\frac{1}{2})$ and $(e_{i1} = -\frac{1}{2}, e_{i2} = \frac{1}{2})$, whereas the extreme values can each occur in

only *one* way. This simple observation reveals a fundamental fact about sums of independent random variables: *intermediate values of sums can occur in more ways than extreme values, and hence tend to be more likely*. It is this property of independent sums that gives rise to their "bell shaped" distributions, as can be seen in parts (c) and (d) of Figure 3.1.

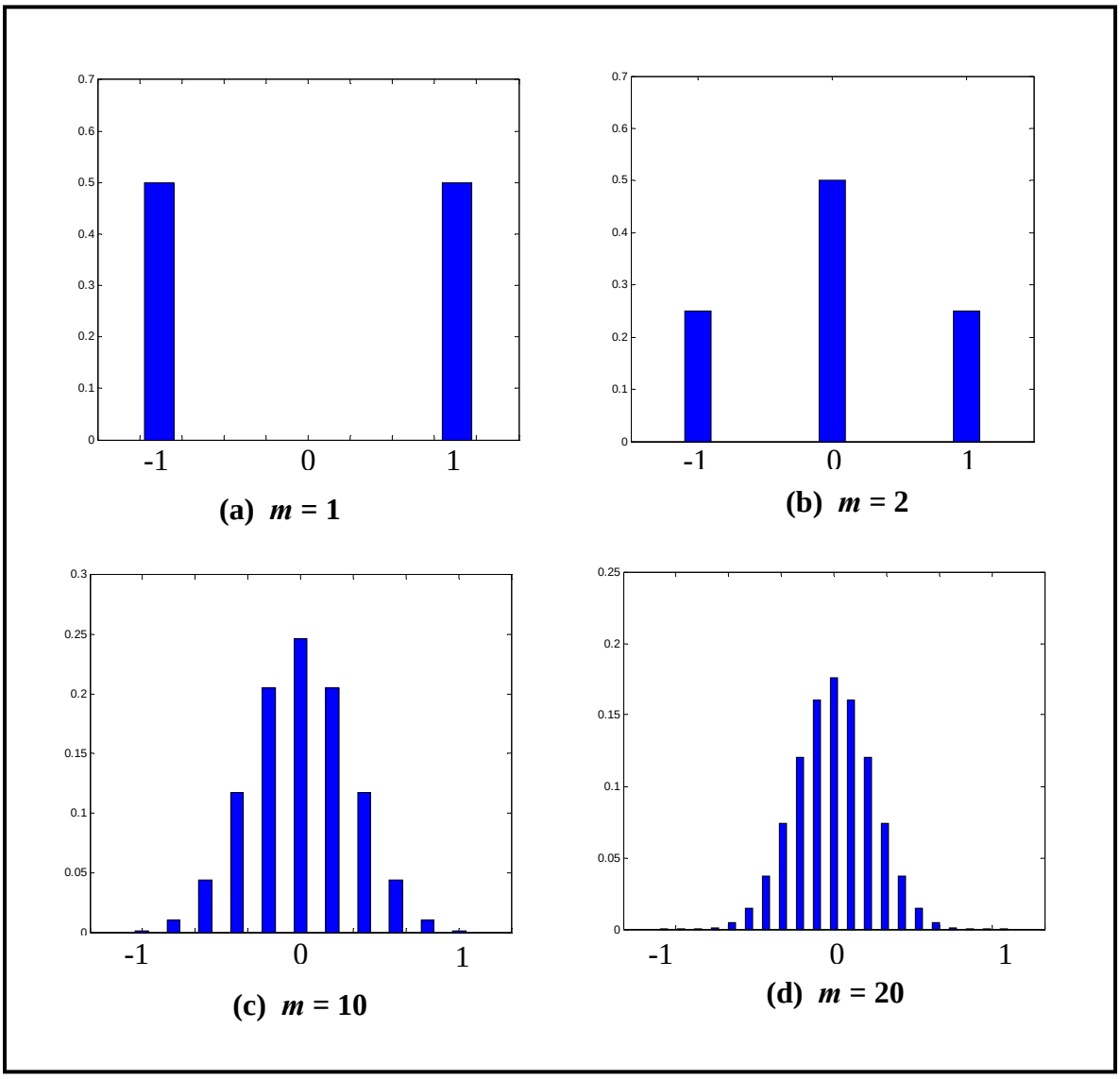


Figure 3.1 Cumulative Binary Errors

But while this basic shape property is easily understood, the truly amazing fact is that the limiting form if this bell shape always corresponds to essentially the *same* distribution, namely the *normal distribution*. To state this precisely, it is important to notice first that

while the distributions in Figure 3.1 start to become bell shaped, they are also starting to concentrate around zero. Indeed, the limiting form of this particular distribution must necessarily be a unit point mass at zero,[1] and is certainly not normally distributed. Here is turns out that the individual values of these factors, $(e_{ik} = 1/m, \text{ or } e_{ik} = -1/m)$, become "too small" as *m* increases, so that eventually even their sum, ε_i, will almost certainly vanish. At the other extreme, suppose that these values are independent of *m*, say $(e_{ik} = 1, \text{ or } e_{ik} = -1)$. Then while these individual values will eventually become small relative to their sum, ε_i, the *variance* of ε_i itself will increase without bound.[2] In a similar manner, observe that if the common *means* of these individual factors were not identically zero, then the limiting mean of ε_i would also be unbounded.[3] So it should be clear that precise analysis of limiting random sums is rather delicate.

3.1.1 Standardized Random Variables

The time-honored solution to these difficulties is to rescale these random sums in a manner which ensures that both their mean and variance remain *constant* as *m* increases. To do so, we begin by observing that for any random variable, X with mean, $\mu = E(X)$, and variance, $\sigma^2 = \text{var}(X)$ the transformed random variable,

$$(3.1.3) \qquad Z = \frac{X - \mu}{\sigma} = \tfrac{1}{\sigma} X - \tfrac{\mu}{\sigma}$$

necessarily has *zero mean* since (by the linearity of expectations),

$$(3.1.4) \qquad E(Z) = \tfrac{1}{\sigma} E(X) - \tfrac{\mu}{\sigma} = \tfrac{\mu}{\sigma} - \tfrac{\mu}{\sigma} = 0$$

Moreover, Z also has *unit variance*, since by (3.4),

$$(3.1.5) \qquad \text{var}(Z) = E(Z^2) = E\left[\left(\frac{X - \mu}{\sigma}\right)^2\right] = E\left[\frac{1}{\sigma^2}(X - \mu)^2\right] = \frac{E[(X - \mu)^2]}{\sigma^2} = 1$$

---

[1] Simply observe that if x_{ik} is a binary random variable with $\Pr(x_{ik} = 1) = .5 = \Pr(x_{ik} = -1)$ then by definition, $e_{ik} = x_{ik}/m$, so that $\varepsilon_i = (x_{i1} + \cdots + x_{im})/m$ is seen to be the *average* of *m* samples from this binary distribution. But by the *Law of Large Numbers*, such sample averages must eventually concentrate at the population mean, $E(x_{ik}) = 0$.

[2] In particular since $\text{var}(e_{ik}) = E(e_{ik}^2) = .5(-1)^2 + .5(-1)^2 = 1$ for all *k*, it would then follow from the independence of individual factors that $\text{var}(\varepsilon_i) = \sum_{k=1}^{m} \text{var}(e_{ik}) = m \cdot \text{var}(e_{1k}) = m$, and hence that $\text{var}(\varepsilon_i) \to \infty$ as $m \to \infty$.

[3] Since $E(\varepsilon_i) = \sum_{k=1}^{m} E(e_{ik}) = m\,E(e_{i1})$ implies $|E(\varepsilon_i)| = m|E(e_{i1})|$, it follows that if $|E(e_{i1})| > 0$ then $|E(\varepsilon_i)| \to \infty$ as $m \to \infty$.

This fundamental transformation procedure is called the *standardization* of *X*. We shall use this device to study the limits of sums. But more generally, it is important to observe that if one wants to compare the distributional "shapes" of any two random variables, say, *X* and *Y*, it is much more convenient to compare their standardizations, Z_X and Z_Y. Since these new variables always have the same mean and variance, a comparison of Z_X and Z_Y thus allows one to focus on qualitative differences in their *shape*.

In particular, we can in principle use this standardization procedure to study the limiting distributional shape of *any* sum of random variables, say

(3.1.6) $$S_m = X_1 + \cdots + X_m = \sum_{k=1}^{m} X_k$$

As in our example, let us assume for the present that these variables are *independently* and *identically distributed* (*iid*), with common *mean*, μ, and *variance*, σ^2 [so that $(X_1,..,X_m)$ can be viewed as a random sample of size *m* from some common distribution]. Then the mean and variance of S_m are given respectively by

(3.1.7) $$E(S_m) = \sum_{k=1}^{m} E(X_k) = \sum_{k=1}^{m} \mu = m\,\mu$$

(3.1.8) $$\text{var}(S_m) = \sum_{k=1}^{m} \text{var}(X_k) = \sum_{k=1}^{m} \sigma^2 = m\,\sigma^2$$

So as above, we may construct the associated *standardized sum*,

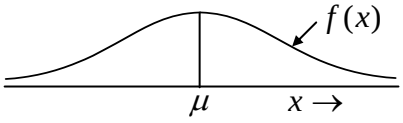
(3.1.9) $$Z_m = \frac{S_m - E(S_m)}{\sqrt{\text{var}(S_n)}} = \frac{S_m - m\mu}{\sqrt{m\sigma^2}}$$

which by definition implies that $E(Z_m) = 0$ and $\text{var}(Z_m) = 1$ for all *m*. The key property of these standardized sums is that for large *m* the distribution of Z_m is approximately normally distributed.

3.1.2 Normal Distribution

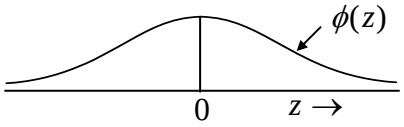
To state this precisely, we must first define the normal distribution. A random variable, *X*, with mean μ and variance σ^2 is said to be *normally distributed*, written, $X \sim N(\mu,\sigma^2)$, if and only if X has probability density given by

(3.1.10) $$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}} = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$



[where the first version shows $f(x)$ as an explicit function of (μ,σ^2) and the second shows the more standard version of $f(x)$ in terms of (μ,σ)]. This is the classical "bell-shaped" curve, centered on the mean, μ, as shown on the right. A key property of normal random variables (that we shall make use of many times) is that *any linear function of a normal random variable is also normally distributed*. In particular, since the standardization procedure in (3.1.3) is seen to be a linear function, it follows that the standardization, Z, of any normal random variable must be normally distributed with mean, $E(Z)=0$, variance, $\text{var}(Z)=1$, and with density

(3.1.11) $$\phi(z)=\frac{1}{\sqrt{2\pi}}\,exp\left(-\frac{z^2}{2}\right)$$



For obvious reasons, this is called the *standard normal distribution* (or *density*), and is generally denoted by ϕ. The importance of this particular distribution is that all probability questions about normal random variables can be essentially answered by standardizing them and applying the standard normal distribution (so that all normal tables are based entirely on this standardized form).

Next, if the *cumulative distribution function* (*cdf*) of any random variable, X, is denoted for all values, x, by $F(x)=Prob(X\le x)$, then for any *standard normal* random variable, $Z\sim N(0,1)$, the *cdf* of Z is denoted by

(3.1.12) $$\Phi(z)\;=\;Prob(Z\le z)\;=\;\int_{-\infty}^{z}\phi(z)\,dz$$

Again Φ is usually reserved for this important *cdf* (that forms the basis of all normal tables).

3.1.3 Central Limit Theorems

With these preliminaries, we can now give a precise statement of the limiting normal property of standardized sums stated above. To do so, it is important to note first that the distribution of any random variable is *completely defined* by its *cdf*. [For example, in the standard normal case above it should be clear that the standard normal distribution, ϕ, is recovered by simply differentiating Φ.] Hence, letting the *cdf* of the standardized sum, Z_m, in (3.1.9) be denoted by F_{Z_m}, we now have the following classical form of the Central Limit Theorem (CLT):

Central Limit Theorem (Classical). *For any sequence of iid random variables* $(X_1,..,X_m)$ *with standardized sum,* Z_m, *in* (3.1.9),

(3.1.13) $$\lim_{m\to\infty} F_{Z_m}(z)\;=\;\Phi(z)\quad \textit{for all } z.$$

In other words, the *cdf* of *iid* standardized sums, Z_m, converges to the *cdf* of the standard normal distribution. The advantage of this *cdf* formulation is that one obtains an *exact limit* result. But in practical terms, the implication of the CLT is that for "sufficiently large" *m*, the distribution of such standardized sums is *approximately* normally distributed.[4] Even more to the point, since (3.1.3) implies that *iid* sums, S_m, are *linear functions* of their standardizations, Z_m, and since linear functions of normal random variables are again normal, it may also be concluded that these sums are *approximately normal*. If for convenience we now use the notation, $X \approx_d N(\mu,\sigma^2)$, to indicate that a random variable *X* is *approximately distributed* normal with mean, μ, and variance, σ^2, and if we recall from (3.1.7) and (3.1.8) that the mean and variance of S_m are given by $m\mu$ and $m\sigma^2$, respectively , then we have the follows more useful form of the CLT :

Central Limit Theorem (Practical). *For all sums, S_m, of iid random variables with m sufficiently large,*

(3.1.14) $$S_m \approx_d N(m\mu, m\sigma^2)$$

This result can in principle be used to motivate the fundamental *normality assumption* about random effects, ε_i. In particular, if ε_i is a sum of *iid* random components as in (3.1.1), with zero means, then by (3.1.14) it follows that ε_i will also be approximately normal with zero mean for sufficiently large *m*.

However, it should be emphasized here that in practical examples (such as the one discussed in Section 3.2 below) the individual components, e_{ik}, of ε_i may not be fully independent, and are of course not likely to be identically distributed. Hence it is important to emphasize that the CLT is actually much more general that the classical assertion above for iid random variables. While such generalizations require conditions that are too technical to even be stated in a precise manner here, [5] it is nonetheless useful to given a very rough statement of the general version as follows: [6]

[4] Recall from footnote 5 in Section 3.2.2 of Part I that "sufficiently large" is usually taken to mean $m \geq 30$, as long as the common distribution of the underlying random variables (X_k) in (3.1.6) is not "too skewed".

[5] For further details about such generalizations, an excellent place to start is the Wikipedia discussion of the CLT at http://en.wikipedia.org/wiki/Central_limit_theorem.

[6] The following version of the Central Limit Theorem (and the multivariate version of this theorem in section 3.2.3 below) based on Theorem 8.11 in Brieman (1969). The advantage of the present version is that it directly extends the "iid" conditions of the classical CLT.

Central Limit Theorem (General). *For any sum,* $S_m = X_1 + \cdots + X_m$, *of random variables with means,* $\mu_1,..,\mu_m$, *and variances,* $\sigma_1^2,..,\sigma_m^2$, *if* (i) *the distributions of these random variables are "not too different", and* (ii) *the dependencies among these random variables is "not too strong", then for sufficiently large m, the distribution of* S_m *is approximately normal, i.e.,*

(3.1.15) $$S_m \approx_d N(\mu, \sigma^2)$$

with $\mu = \mu_1 + \cdots + \mu_m$ *and* $\sigma^2 = \sigma_1^2 + \cdots + \sigma_m^2$.

So for random effects, $\varepsilon_i = e_{i1} + \cdots + e_{im}$, with total variance, $\sigma^2 = \sigma_1^2 + \cdots + \sigma_m^2$, it follows that as long as conditions (i) and (ii) are reasonable and *m* is sufficiently large, random effects, ε_i, will be approximately normally distributed as

(3.1.16) $$\varepsilon_i \approx_d N(0, \sigma^2)$$

3.1.4 CLT for the Sample Mean

While the main application of the CLT for our present purposes is to motivate the normality assumption about residuals in a host of statistical models (including linear regression), it is important to add that perhaps the single most important application of the CLT is for inference about population means. In particular, if one draws a *iid* random sample, $(X_1,..,X_m)$ from a population with unknown mean, μ, and constructs the associated sample mean:

(3.1.17) $$\bar{X}_m = \tfrac{1}{m}\sum_{k=1}^{m} X_k = \tfrac{1}{m} S_m \quad ,$$

then by (3.1.7) the identity,

(3.1.18) $$E(\bar{X}_m) = \tfrac{1}{m} E(S_m) = \tfrac{1}{m}(m\mu) = \mu$$

implies that $\bar{X}_m$ is the natural *unbiased estimator* of μ. Moreover, by (3.1.8), the second identity,

(3.1.19) $$\text{var}(\bar{X}_m) = \tfrac{1}{m^2}\text{var}(S_m) = \tfrac{1}{m^2}(m\sigma^2) = \sigma^2/m$$

implies that for large *m* this estimate has a small variance, and hence should be close to μ (which is of course precisely the Law of Large Numbers). But one can say even more by the CLT. To do so, note first that the *standardized sample mean*,

(3.1.20) $$Z_{\bar{X}_m} \equiv \frac{\bar{X}_m - E(\bar{X}_m)}{\sigma(\bar{X}_m)} = \frac{\bar{X}_m - \mu}{\sqrt{\sigma^2/m}}$$

can equivalently be written as

(3.1.21) $$Z_{\bar{X}_m} = \frac{\frac{1}{m}S_m - \mu}{\sqrt{\sigma^2/m}} = \frac{S_m - m\mu}{m\sqrt{\sigma^2/m}} = \frac{S_m - m\mu}{\sqrt{m\sigma^2}} = Z_m$$

and hence satisfies exactly the same limiting properties as the sample sum. In particular this yields the follows version of the practical CLT in (3.1.14) above for sample means:

> **Central Limit Theorem (Sample Means).** *For sufficiently large iid random samples,* $(X_1,..,X_m)$, *from any given statistical population with mean,* μ, *and variance,* σ^2, *the sample mean,* $\bar{X}_m$, *is approximately normal, i.e.,*
>
> (3.1.22) $$\bar{X}_m \approx_d N(\mu, \sigma^2/m)$$

Note in particular that random samples from the same population are by definition *identically* distributed. So as long as they are also *independent*, Corollary 2 is always applicable. But the Clark-Evans test in Section 3.2.2 of Part I provides a classic example where this latter assumption may fail to hold. More generally, the types of dependencies inherent in spatial (or temporal) data require more careful analysis when applying the CLT to sample means.

3.2 Multi-Location Random Effects

Given the above results for random effects, ε_i at individual locations, s_i, we now consider the *vector*, ε, of such random effects for a given set of sample locations, $\{s_i : i = 1,..,n\} \subset R$, i.e.,

(3.2.1) $$\varepsilon = (\varepsilon_i : i = 1,..,n)' = [\varepsilon(s_i) : i = 1,..,n]'$$

As a parallel to (3.1.1) we again assume that these random effects are the cumulative sum of independent factors,

(3.2.2) $$\varepsilon = e_1 + e_2 + \cdots + e_m = \sum_{k=1}^{m} e_k$$

where by definition each independent factor, e_k, is itself a random vector over sample locations, i.e.,

(3.2.3) $$e_k = (e_{ik} : i = 1,..,n)' = [e_k(s_i) : i = 1,..,n]'$$

As one illustration, recall the California rainfall example in which annual precipitation, Y_i, at each of the $n = 30$ sample locations in California was assumed to depend on four explanatory variables ($x_{i1} =$ "altitude", $x_{i2} =$ "latitude", $x_{i2} =$ "distance to coast", and $x_{i4} =$ "rain shadow", as follows

$$(3.2.4) \qquad Y_i = \beta_0 + \sum_{j=1}^{4} \beta_j x_{ij} + \varepsilon_i \ , \quad i = 1,..,n$$

Here the unobserved residuals, ε_i, are the random effects we wish to model. If we write (3.2.4) in vector form as

$$(3.2.5) \qquad Y = \beta_0 1_n + \sum_{j=1}^{4} \beta_j x_j + \varepsilon$$

[where $1_n = (1,..,1)'$ is the unit column vector], then the *residual vector*, ε, in (3.2.5) is an instance of (3.2.1) with $n = 30$. This random vector by definition contains all factors influencing precipitation other that the four "main" effects posited above. So the key assumption in (3.2.2) is that the influence of each unobserved factor is only a small additive part of the total residual effect, ε, not accounted for by the four main effects above.

For example, the first factor, e_1, might be a "cloud cover" effect. More specifically, the unobserved value, $e_{1i} = e_1(s_i)$ at each location, s_i, might represent fluctuations in cloud cover at s_i [where higher (lower) levels of cloud cover tend to contribute positively (negatively) to precipitation at s_i]. Similarly, factor e_2 might be an "atmospheric pressure" effect, where $e_{2i} = e_2(s_i)$ now represents fluctuations in barometric pressure levels at s_i [and where in this case higher (lower) pressure levels tend to contribute negatively (positively) to precipitation levels].

The key point to observe is that while fluctuations in factors like cloud cover or atmospheric pressure will surely exhibit strong spatial dependencies, the dependency *between* these factors at any given location is much weaker. In the present instance, while there may indeed be some degree of negative relation between fluctuations in pressure and cloudiness (e_{1i}, e_{2i}) at any given location, s_i, this tends to be much weaker than the positive relations between either fluctuations in cloud cover (e_{1i}, e_{1j}), or atmospheric pressure (e_{2i}, e_{2j}), at locations, s_i and s_j, that are in close proximity. Hence while the random vectors, e_1 and e_2, can each exhibit strong internal spatial dependencies, it is not unreasonable to treat them as mutually independent. More generally, as a parallel to section (3.1.3) above, it will turn out that if (i) the individual distributions of the random component vectors, $e_1,..,e_m$, in (3.2.2) are not "too different", and (ii) the statistical dependencies between these components are not "too strong", then their sum, ε, will be approximately "normal" for *m* sufficiently large.

But in order to make sense of this statement, we must first extend the normal distribution in (3.1.10) to its multivariate version. This is done in the next section, where we also develop its corresponding invariance property under linear transformations. This will be followed by a development of the multivariate version of the Central Limit Theorem that underscores the importance of this distribution.

3.2.1 Multivariate Normal Distribution

To motivate the multivariate normal (or multi-normal) distribution observe that there is one case in which we can determine the joint distribution of a random vector, $X = (X_1,..,X_n)'$, in terms of the marginal distributions of its component, $X_1,..,X_n$, namely when these components are *independently distributed.* In particular, suppose that each X_i is independently normally distributed as in (3.1.10) with density

$$(3.2.6) \qquad f_i(x_i) = \frac{1}{\sqrt{2\pi\sigma_i^2}}\, e^{-\frac{(x_i-\mu_i)^2}{2\sigma_i^2}} \;,\; i = 1,..,n$$

Then letting, $\sigma_{ii} = \sigma_i^2$ and using the exponent notation, $\sqrt{a} = a^{1/2}$, it follows that the joint density, $f(x_1,..,x_n)$, of X is given by the *product* of these marginals, i.e.,

$$(3.2.7) \qquad f(x_1,..,x_n) = f_1(x_1)\, f_2(x_2)\cdots f_n(x_n)$$

$$= \left(\frac{1}{\sqrt{2\pi\sigma_{11}}}\, e^{-\frac{(x_i-\mu_i)^2}{2\sigma_{11}}}\right)\left(\frac{1}{\sqrt{2\pi\sigma_{22}}}\, e^{-\frac{(x_i-\mu_i)^2}{2\sigma_{22}}}\right)\cdots\left(\frac{1}{\sqrt{2\pi\sigma_{nn}}}\, e^{-\frac{(x_i-\mu_i)^2}{2\sigma_{nn}}}\right)$$

$$= (2\pi)^{-n/2}(\sigma_{11}\sigma_{22}\cdots\sigma_{nn})^{-1/2}\, e^{-\frac{1}{2}\left[\frac{(x_i-\mu_i)^2}{\sigma_{11}}+\cdots+\frac{(x_n-\mu_n)^2}{\sigma_{nn}}\right]}$$

where the last line uses the identity, $(e^{a_1})(e^{a_2})\cdots(e^{a_n}) = e^{a_1+a_2+\cdots+a_n}$. To write this in matrix form, observe first that if $x = (x_1,..,x_n)'$ now denotes a typical realization of random vector, $X = (X_1,..,X_n)'$, then by (3.2.6) the associated mean vector of X is given by $\mu = (\mu_1,..,\mu_n)'$ [as in expression (1.1.4)]. Moreover, since independence implies that $\text{cov}(X_i, X_j) = \sigma_{ij} = 0$ for $i \neq j$, it follows that the covariance matrix of X now takes the form [as in expression (1.1.7)],

(3.2.7) $$\text{cov}(X) = \Sigma = \begin{pmatrix} \sigma_{11} & & & \\ & \sigma_{22} & & \\ & & \ddots & \\ & & & \sigma_{nn} \end{pmatrix}$$

But since the inverse of a diagonal matrix is simply the diagonal matrix of inverse values,

(3.2.8) $$\Sigma^{-1} = \begin{pmatrix} \sigma_{11}^{-1} & & & \\ & \sigma_{22}^{-1} & & \\ & & \ddots & \\ & & & \sigma_{nn}^{-1} \end{pmatrix}$$

it follows that

(3.2.9) $$(x-\mu)'\Sigma^{-1}(x-\mu) =$$

$$(x_1-\mu_1, x_2-\mu_2, .., x_n-\mu_n)\begin{pmatrix} \sigma_{11}^{-1} & & & \\ & \sigma_{22}^{-1} & & \\ & & \ddots & \\ & & & \sigma_{nn}^{-1} \end{pmatrix}\begin{pmatrix} x_1-\mu_1 \\ x_2-\mu_2 \\ \vdots \\ x_n-\mu_n \end{pmatrix}$$

$$= (x_1-\mu_1, x_2-\mu_2, .., x_n-\mu_n)\begin{bmatrix} (x_1-\mu_1)/\sigma_{11} \\ (x_2-\mu_2)/\sigma_{22} \\ \vdots \\ (x_n-\mu_n)/\sigma_{nn} \end{bmatrix}$$

$$= \frac{(x_1-\mu_1)^2}{\sigma_{11}} + \frac{(x_2-\mu_2)^2}{\sigma_{22}} + \cdots + \frac{(x_n-\mu_n)^2}{\sigma_{nn}}$$

which is precisely the exponent sum in (3.2.7). Finally, since the *determinant*, $|\Sigma|$, of a diagonal matrix, Σ, is simply the product of its diagonal elements, i.e.,

(3.2.10) $$|\Sigma| = \begin{vmatrix} \sigma_{11} & & & \\ & \sigma_{22} & & \\ & & \ddots & \\ & & & \sigma_{nn} \end{vmatrix} = \sigma_{11}\sigma_{22}\cdots\sigma_{nn} \ ,$$

we see from (3.2.9) and (3.2.10) that (3.2.7) can be rewritten in matrix form as

(3.2.11) $$\boxed{f(x) = (2\pi)^{-n/2}|\Sigma|^{-1/2} e^{-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)}}$$

This is in fact an instance of the *multi-normal density* (or *multivariate normal density*). More generally, a random vector, $X = (X_1,..,X_n)'$, with associated mean vector, $\mu = (\mu_1,..,\mu_n)'$, and covariance matrix, $\Sigma = (\sigma_{ij} : i, j = 1,..,n)$, is said to be *multi-normally distributed* if and only if its joint density is of the form (3.2.11) for this choice of μ and Σ. As a generalization of the univariate case, this is denoted symbolically by $X \sim N(\mu,\Sigma)$.

While it is not possible to visualize this distribution in high dimensions, we can gain some insight by focusing on the 2-dimensional case, known as the *bi-normal* (or *bivariate normal*) *distribution*. If $X = (X_1, X_2)$ is bi-normally distributed with mean vector, $\mu = (\mu_1, \mu_2)'$ and covariance matrix,

$$(3.2.12) \qquad \Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix}$$

then the basic shape of the density function in (3.2.11) is largely determined by the *correlation* between X_1 and X_2, i.e., by

$$(3.2.13) \qquad \rho(X_1, X_2) = \frac{\text{cov}(X_1, X_2)}{\sigma(X_1)\sigma(X_2)} = \frac{\sigma_{12}}{\sqrt{\sigma_{11}}\sqrt{\sigma_{22}}}$$

This is most easily illustrated by setting $\mu_1 = \mu_2 = 0$ and $\sigma_{11} = \sigma_{22} = 1$ so that the only parameter of this distribution is covariance, σ_{12}, which in this case is seen from (3.2.13) to be precisely the *correlation*, ρ, between X_1 and X_2. The independence case $(\rho = 0)$ is shown in Figure 3.2 below, which is simply a 2-dimensional version of the standard normal distribution in (3.1.11) above. Indeed both of its marginal distributions are identical with (3.1.11). Figure 3.3 depicts a case with extreme positive correlation $(\rho = .8)$ to emphasize the role of correlation in shaping this distribution. In particular, this high correlation implies that value pairs (x_1, x_2) that are similar in magnitude (close to the $45°$ line) are more likely to occur, and hence have higher probability density. Thus the density is more concentrated along the $45°$ line, as shown in the figure.

These properties persist in higher dimensions as well. In particular, the "bell-shaped" concentration of density around the origin continues to hold in higher dimensions, and is more elongated in those directions where correlations between components are more extreme.

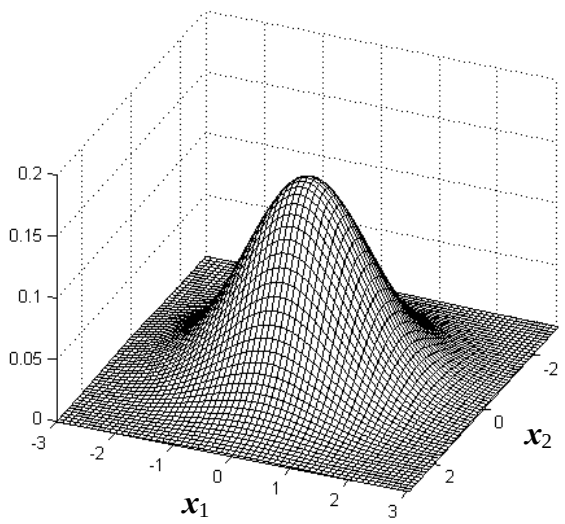


Figure 3.2. Bi-normal Distribution (ρ = 0)

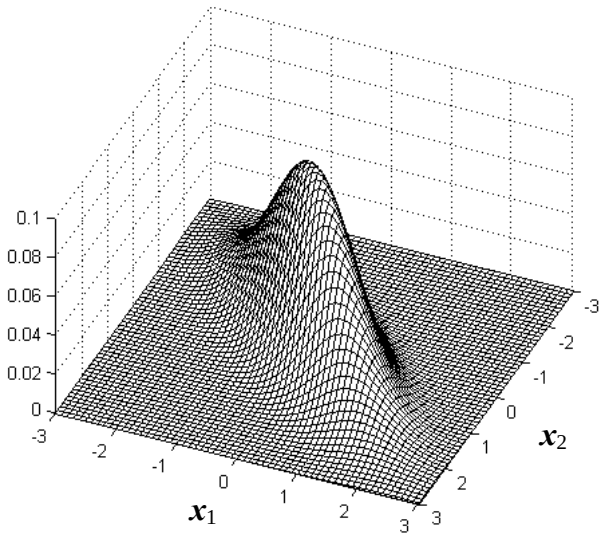


Figure 3.3. Bi-normal Distribution (ρ = .8)

3.2.2 Linear Invariance Property

For purposes of analysis, the single most useful feature of this distribution is that *all linear transformations of multi-normal random vectors are again multi-normal*. To state this precisely, we begin by calculating the mean and covariance matrix for general linear transformations of random vectors. Given a random vector, $X = (X_1,..,X_n)'$, with mean vector, $E(X) = \mu = (\mu_1,..,\mu_n)'$ and covariance matrix, $\text{cov}(X) = \Sigma$, together with any compatible $(m \times n)$ matrix, $A = (a_{ij} : i = 1,..,m,\ j = 1,..,n)$, and n-vector, $b = (b_1,..,b_n)'$ of coefficients, consider the *linear transformation* of X defined by

(3.2.14) $\quad Y = AX + b$

Following standard conventions, if $m = 1$ then the $(1 \times n)$ matrix, A, is usually written as the transpose of an *n*-vector, $a' = (a_1,..,a_n)$, so that (3.2.14) takes the form,

(3.2.15) $\quad Y = a'X + b$

where b is a scalar. If $b = 0$ then the random variable, $Y = a'X$, is called a *linear compound* of X. For example, each component of X can be identified by such a linear compound as follows. If the columns of the n-square *identity matrix*, I_n, are denoted by

(3.2.16) $$I_n = \begin{bmatrix} 1 & & \\ & 1 & \\ & & \ddots & \\ & & & 1 \end{bmatrix} = \left[\begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, ..., \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix} \right] = [e_1, e_2, ..., e_n]$$

then by setting $a = e_i$ and $b = 0$ in (3.2.15), we see that

(3.2.17) $$X_i = e_i' X \ , \quad i = 1,..,n$$

So linear transformations provide a very flexible tool for analyzing random vectors.

Next recall from the linearity of expectations that by taking expectations in (3.2.14) we obtain

(3.2.18) $$E(Y) = E(AX + b) = AE(X) + b = A\mu + b$$

By using this result, we can obtain the covariance matrix for Y as follows. First note that by definition the expected value of a matrix of random variable is simply the matrix of their expectations, i.e.,

(3.2.19) $$E\begin{pmatrix} Z_{11} & \cdots & Z_{1n} \\ \vdots & \ddots & \vdots \\ Z_{m1} & \cdots & Z_{mn} \end{pmatrix} = \begin{pmatrix} E(Z_{11}) & \cdots & E(Z_{1n}) \\ \vdots & \ddots & \vdots \\ E(Z_{m1}) & \cdots & (Z_{mn}) \end{pmatrix}$$

So the definition of $\text{cov}(Y)$ in (1.1.7) can equivalently be written in matrix terms as

(3.2.20) $$\text{cov}(Y) = \begin{pmatrix} E[(Y_1 - \mu_1)(Y_1 - \mu_1)] & \cdots & E[(Y_1 - \mu_1)(Y_n - \mu_n)] \\ \vdots & \ddots & \vdots \\ E[(Y_n - \mu_n)(Y_1 - \mu_1)] & \cdots & E[(Y_n - \mu_n)(Y_n - \mu_n)] \end{pmatrix}$$

$$= E\begin{pmatrix} (Y_1 - \mu_1)(Y_1 - \mu_1) & \cdots & (Y_1 - \mu_1)(Y_n - \mu_n) \\ \vdots & \ddots & \vdots \\ (Y_n - \mu_n)(Y_1 - \mu_1) & \cdots & (Y_n - \mu_n)(Y_n - \mu_n) \end{pmatrix}$$

$$= E\left[\begin{pmatrix} Y_1 - \mu_1 \\ \vdots \\ Y_n - \mu_n \end{pmatrix} (Y_1 - \mu_1, ..., Y_n - \mu_n) \right]$$

$$= E[(Y - \mu)(Y - \mu)']$$

By applying this to (3.2.15) we obtain the following very useful result:

$$
\begin{aligned}
(3.2.21)\quad \mathrm{cov}(Y) &= E[(Y-\mu)(Y-\mu)'] \\
&= E\{([AX+b]-[A\mu+b])([AX+b]-[A\mu+b])'\} \\
&= E[(AX-A\mu)(AX-A\mu)'] \\
&= E[A(X-\mu)(X-\mu)'A'] \\
&= AE[(X-\mu)(X-\mu)']A' \\
&= A\,\mathrm{cov}(X)\,A' \\
&\Rightarrow \boxed{\mathrm{cov}(AX) \;=\; A\,\Sigma\,A'}
\end{aligned}
$$

So both the mean and covariance matrix of $AX+b$ are directly obtainable from those of X. We shall use these properties many times in analyzing the multivariate spatial models of subsequent sections.

But for the moment, the key feature of these results is that the distribution of any linear transformation, $AX+b$, of a multi-normal random vector, $X \sim N(\mu,\Sigma)$, is obtained by simply replacing the mean and covariance matrix of X in (3.2.11) with those of $AX+b$. The only requirement here is that the resulting covariance matrix, $A\Sigma A'$, be *nonsingular* so that the inverse covariance matrix, $(A\Sigma A')^{-1}$, in (3.2.11) exists. This in turn is equivalent to the condition that the rows of A be linearly independent vectors, so that A is said to be of *full row rank*. With this stipulation, we have the following result [established in Section A3.2.3 of the Appendix to Part III in this NOTEBOOK]:[7]

> **Linear Invariance Theorem**. *For any multi-normal random vector, $X \sim N(\mu,\Sigma)$, and linear transformation, $Y = AX+b$, of X with A of full row rank, Y is also multi-normally distributed as*
>
> (3.2.22) $\quad Y \sim N(A\mu+b, A\Sigma A')$

What this means in practical terms is that if a given random vector, X, is known (or assumed) to be multi-normally distributed as $X \sim N(\mu,\Sigma)$, then we can immediately write down the *exact distribution* of essentially any linear function, $AX+b$, of X.

3.2.3 Multivariate Central Limit Theorem

We are now ready to consider multivariate extensions of the univariate central limit theorems above. Our objective here is to develop only those aspects of the multivariate

[7] For an alternative development of this important result, see for example Theorem 2.4.4 in Anderson (1958).

case that are relevant for our present purposes. The first objective is to show that the multivariate case relates to the univariate case in a remarkably simple way. To do so, recall first from (3.2.17) above that for any random vector, $X = (X_1,..,X_n)$, each of its components, X_i, can be represented as a linear transformation, $X_i = e_i'X$, of X. So each marginal distribution of X is automatically the distribution of this linear compound. More generally, each linear compound, $a'X$, can be said to define a *generalized marginal distribution* of X.[8] Now while the marginal distributions of X only determine its joint distribution in the case of independence [as in (3.2.7) above], it turns out that the joint distribution of X is always *completely determined by its generalized marginal distributions*.[9] To appreciate the power of this result, recall from the Linear Invariance Theorem above that if X is multi-normal with mean vector, μ, and covariance matrix, Σ, then all of its linear compounds, $a'X$, are automatically univariate normally distributed with means, $a'\mu$, and variances, $a'\Sigma a$. But since these marginals in turn uniquely determine the distribution of X, it must necessarily be *multi-normal*. Thus we are led to the following fundamental correspondence:

Univariate-Multivariate Correspondence. *A random vector, X, with mean vector, μ, and covariance matrix, Σ, is multi-normally distributed as*

(3.2.23) $$X \sim N(\mu,\Sigma)$$

if and only if every linear compound, $a'X$, is univariate normal, i.e.,

(3.2.24) $$a'X \sim N(a'\mu, a'\Sigma a)$$

In view of this correspondence, it is not surprising that there is an intimate relation between univariate and multivariate central limit theorems. In particular, if any of the univariate conditions in the central limit theorems above hold for all generalized marginal distributions of X, then X will automatically be asymptotically multivariate normal. For example, if as an extension of (3.1.15) one considers a sum of *iid* random vectors,

(3.2.25) $$S_m = X_1 + \cdots + X_m$$

then it follows at once that the terms in each linear compound,

(3.2.26) $$a'S_m = a'X_1 + \cdots + a'X_m$$

must necessarily be *iid* as well. Hence we obtain an immediate extension of the "Practical" Central Limit Theorem in (3.1.14) above

[8] Since each marginal compound, $e_i'X$, has a coefficient vector of unit length, i.e., $\| e_i \| = 1$, it is formally more appropriate to restrict generalized marginals to linear compounds, a, of unit length ($\| a \| = 1$). But for our present purposes we need not be concerned with such scaling effects.

[9] For a development of this idea (due to Cramer and Wold), see Theorem 29.4 in Billingsley (1979).

Multivariate Central Limit Theorem (Practical). *For all sums of iid random vectors,* $S_m = X_1 + \cdots + X_m$, *with common mean vector,* μ, *and covariance matrix,* Σ, *if m sufficiently large then*

(3.2.27) $$S_m \approx_d N(m\mu, m\Sigma)$$

But since multivariate normality will almost always arise as a *model assumption* in our spatial applications, the most useful extension is the "General" Central Limit Theorem in (3.1.15), which may now be stated as follows:[10]

Multivariate Central Limit Theorem (General). *For any sum,* $S_m = X_1 + \cdots + X_m$, *of random vectors with individual means,* $\mu_1, .., \mu_m$, *and covariance matrices,* $\Sigma_1, .., \Sigma_m$, *if* (i) *the distributions of these random vectors are "not too different", and* (ii) *the dependencies among these random vectors are "not too strong", then for sufficiently large m, the distribution of* S_m *is approximately multi-normal, i.e.,*

(3.2.28) $$S_m \approx_d N(\mu, \Sigma)$$

with $\mu = \mu_1 + \cdots + \mu_m$ *and* $\Sigma = \Sigma_1 + \cdots + \Sigma_m$.

Finally, it is appropriate to restate this result explicitly in terms of *multi-location random effects*, which form the central focus of this section.

Spatial Random Effects Theorem. *For any random vector of multi-location effects,* $\varepsilon = (\varepsilon_i : i = 1, .., n)'$, *comprised of a sum of individual random factors,* $\varepsilon = e_1 + e_2 + \cdots + e_m$, *with zero means and covariance matrices,* $\Sigma_1, .., \Sigma_m$, *if* (i) *the distributions of these random factors are "not too different", and* (ii) *the dependencies among these random factors are "not too strong", then for sufficiently large m, the distribution of* ε *is approximately multi-normal, i.e.,*

(3.2.29) $$\varepsilon \approx_d N(0, \Sigma)$$

with $\Sigma = \Sigma_1 + \cdots + \Sigma_m$.

It is this version of the Central Limit Theorem that will form the basis for essentially all random-effects models in the analyses to follow.

[10] For a similar (informal) statement of this general version of the Multivariate Central Theorem, see Theorem 8.11 in Brieman (1969).

3.3 Spatial Stationarity

Given the *Spatial Random Effects Theorem* above, the task remaining is to specify the unknown covariance matrix, Σ, for these random effects. Since Σ is in turn a sum of individual covariance matrices, Σ_k, for random factors $k = 1,...,m$, it might seem better to specify these individual covariance structures. But rather than attempt to identify such factors, our strategy will be to focus on general spatial dependencies that should be common to all these covariance structures, and hence should be exhibited by Σ. In doing so, it is also important to emphasize that such statistical dependencies often have little substantive relation to the main phenomena of interest. In terms of our basic modeling framework, $Y(s) = \mu(s) + \varepsilon(s)$, in (1.2.1) above, we are usually much more interested in the global structure of the spatial process, as represented by $\mu(s)$, than in the specific relations among unobserved residuals $\{\varepsilon(s_i) : i = 1,..,n\}$ at sample locations $\{s_i : i = 1,..,n\}$. Indeed, these relations are typically regarded as "second-order" effects in contrast to the "first-order" effects represented by $\mu(s)$. Hence it is desirable to model such second-order effects in a manner that will allow the analysis to focus on the first-order effects, while at the same time taking these unobserved dependencies into account. This general strategy can be illustrated by the following example.

3.3.1 Example: Measuring Ocean Depths

Suppose that one is interested in mapping the *depth* of the sea floor over a given region. Typically this is done by taking echo soundings (sonar measurements) at regular intervals from a vessel traversing a system of paths over the ocean surface. This will yield a set of *depth readings*, $\{D_i = D(s_i) : i = 1,..,n\}$, such as the set of measurements is shown in Figure 3.4 below:

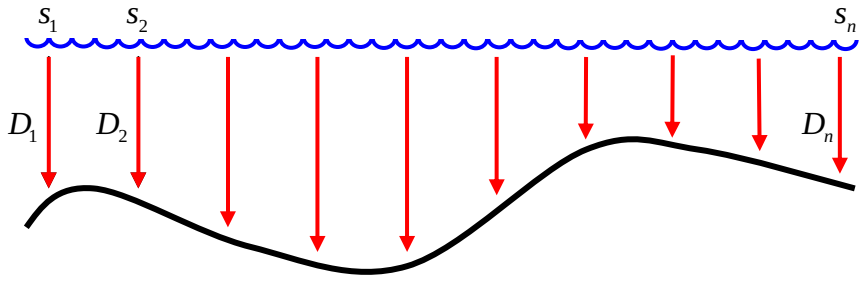


Figure 3.4. Pattern of Depth Measurements

However, the ocean is not a homogeneous medium. In particular, it is well known that such echo soundings can be influenced by the local concentration of zooplankton in the region of each sounding. These clouds of zooplankton (illustrated in Figure 3.5 below) create interference called "ocean volume reverberation".

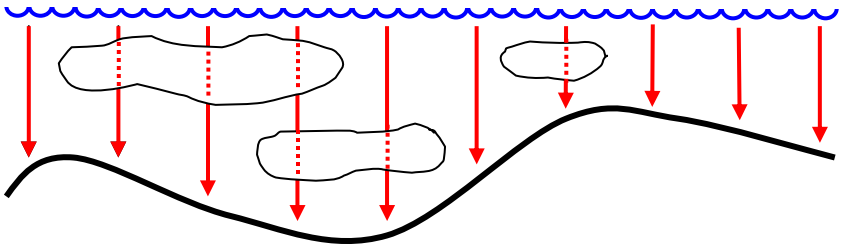


Figure 3.5. Zooplankton Interference

These interference patterns tend to vary from location to location, and even from day to day (much in the same way that sunlight is affected by cloud patterns).[11] So actual readings are random variables of the form,

(3.3.1) $$D(s_i) = d(s_i) + \varepsilon(s_i) \ , \ i = 1,..,n$$

where in this case the actual depth at location s_i is represented by $d(s_i) = E[D(s_i)]$, and $\varepsilon(s_i)$ represents measurement error due to interference.[12] Moreover these errors are *statistically dependent*, since plankton concentrations at nearby locations will tend to be more similar than at locations widely separated in space. Hence to obtain confidence bounds on the true depth at location s_i, it is necessary to postulate a statistical model of these joint interference levels, $[\varepsilon(s_i) : i = 1,..,n]$. Now one could in principle develop a detailed model of zooplankton behavior, including their patterns of individual movement and clustering behavior. However, such models are not only highly complex in nature, they are very far removed from the present target of interest, which is to obtain accurate depth measurements.[13]

[11] Actual variations in the distribution of zooplankton are more diffuse than the "clouds" depicted in Figure 3.5. Vertical movement of zooplankton in the water column is governed mainly by changes in sunlight, and horizontal movement by ocean currents.

[12] In actuality, such measurement errors include many different sources, such as the reflective properties of the sea floor. Moreover, depth measurements are actually made indirectly in terms of the *transmission loss*, $L_i = L(s_i)$, between the signal sent and the echo received. The corresponding depth, D_i, is obtained from L_i by a functional relation, $D_i = \phi(L_i, \theta)$, where θ is a vector of parameters that have been calibrated under "idealized" conditions. For further details, see Urick, R.J. (1983) *Principles of Underwater Sound*, 3rd ed., McGraw-Hill: New York, and in particular the discussion around p.413.

[13] Here it important to note that such detailed models can be of great interest in other contexts. For example, acoustic signals are also used to estimate the volume of zooplankton available as a food source for sea creatures higher in the food chain. To do so, it is essential to relate acoustic signals to the detailed behavior of such microscopic creatures. See for example, Stanton T.K. and D. Chu (2000) "Review and recommendations for the modeling of acoustic scattering by fluid-like elongated zooplankton: euphausiids and copepods", *ICES Journal of Marine Science*, 57: 793–807.

So what is needed here is a statistical model of spatial residuals that allows for *local spatial dependencies*, but is simple enough to be estimated explicitly. To do so, we will adopt the following basic assumptions of *spatial stationarity*:

(3.3.2) [*Homogeneity*] Residuals, $\varepsilon(s_i)$, are *identically distributed* at all locations s_i.

(3.3.3) [*Isotropy*] The joint distribution of distinct residuals, $\varepsilon(s_i)$ and $\varepsilon(s_j)$ depends only on the *distance* between locations s_i and s_j.

These assumptions are loosely related to the notion of "isotropic stationarity" for point processes discussed in Section 2.5 of Part I. But here we focus on the joint distribution of random variables at selected locations in space rather than point counts in selected regions of space. To motivate the present assumptions in the context of our example, observe first that while zooplankton concentrations at any point of time may differ between locations, it can be expected that the range of possible concentration levels over time will be quite similar at each location. More generally, the *Homogeneity* assumption asserts that the *marginal distributions* of these concentration levels are the same at each location. To appreciate the need for such an assumption, observe first that while it is in principle possible to take many depth measurements at each location and employ these samples to estimate location-specific distributions of each random variable, this is generally very costly (or even infeasible). Moreover, the same is true of most spatial data sets, such as the set of total rainfall levels or peak daily temperatures reported by regional weather stations on a given day. So in terms of the present example, one typically has a *single* set of depth measurements $[D(s_i): i = 1,..,n]$, and hence only a *single joint realization* of the set of unobserved residuals $[\varepsilon(s_i): i = 1,..,n]$. Thus, without further assumptions, it is impossible to say *anything* statistically about these residuals. From this viewpoint, the fundamental role of the Homogeneity assumption is to allow the joint realizations, $[\varepsilon(s_i): i = 1,..,n]$, to be treated as multiple samples from a *common population* that can be used to estimate parameters of this population.

The *Isotropy* assumption is very similar in spirit. But here the focus is on statistical dependencies between distinct random variables, $\varepsilon(s_i)$ and $\varepsilon(s_j)$. For even if their marginal distributions are known, one cannot hope to say anything further about their *joint* distribution on the basis of a single sample. But in the present example it is reasonable to assume that if a given cloud of zooplankton (in Figure 3.5) covers location, s_i, then it is very likely to cover locations s_j which are sufficiently close to s_j. Similarly for locations that are very far apart, it is reasonable to suppose that clouds covering s_i have little to do with those covering s_j. Hence the Isotropy assumption asserts more generally that similarities between concentration levels at different locations depend only on the distance between them. The practical implication of this assumption is that all

pairs of residuals, $\varepsilon(s_i)$ and $\varepsilon(s_j)$, separated by the same distance, $h = \|s_i - s_j\|$, must exhibit the *same* degree of dependency. Thus a collection of such pairs can in principle provide multiple samples to estimate the degree of statistical dependency at any given distance, h. A second advantage of this Isotropy assumption is that it allows simple models of "local spatial dependency" to be formulated directly in terms of this single distance parameter. So it should be clear that these two assumptions of spatial stationarity do indeed provide a natural starting point for the desired statistical model of residuals.

But before proceeding, it should also be emphasized that while these assumptions are conceptually appealing and analytically useful – they may of course be *wrong*. For example, it can be argued in the present illustration that locations in shallow depths (Figure 3.5) will tend to experience lower concentration levels than locations in deeper waters. If so, then the Homogeneity assumption will fail to hold. Hence more complex models involved "nonhomogeneous" residuals may be required in some cases.[14] As a second example, suppose that the spatial movement of zooplankton is known to be largely governed by prevailing ocean currents, so that clouds of zooplankton tend to be more elongated in the direction of the current. If so, then spatial dependencies will depend on direction as well as distance, and the Isotropy assumption will fail to hold. Such cases may require more complex "anisotropic" models of spatial dependencies.[15]

3.3.2. Covariance Stationarity

In many cases the assumptions above are stronger than necessary. In particular, recall from the *Spatial Random Effects Theorem* (together with the introductory discussion in Section 3.3) that such random effects are already postulated to be *multi-normally* distributed with zero means. So all that is required for our purposes is that these homogeneity and isotropy assumptions be reflected by the matrix, Σ, of covariances among these random effects.

To do so, it will be convenient for our later purposes to formulate such covariance properties in terms of more general spatial stochastic processes. A spatial stochastic process, $\{Y(s) : s \in R\}$, is said to be *covariance stationary* if and only if the following two conditions hold for all $s_1, s_2, v_1, v_2 \in R$:

$$E[Y(s_1)] = E[Y(s_2)] \tag{3.3.4}$$

$$\| s_1 - s_2 \| = \| v_1 - v_2 \| \Rightarrow \operatorname{cov}[Y(s_1), Y(s_2)] = \operatorname{cov}[Y(v_1), Y(v_2)] \tag{3.3.5}$$

These conditions can be stated more compactly by observing that (3.3.4) implies the existence of a *common mean value*, μ, for all random variables. Moreover, (3.3.5)

[14] For example, it might be postulated that the variance of $\varepsilon(s)$ depends on the unknown true depth, $d(s)$, at each location, s. Such nonstationary formulations are complex, and beyond the scope of these notes.

[15] Such models are discussed for example by Gotway and Waller (2004, Section 2.8.5).

implies that covariance depends only on distance, so that for each distance, h, and pair of locations $s,v \in R$ with $\|s-v\| = h$ there exists a *common covariance value*, $C(h)$, such that $\text{cov}[Y(s),Y(v)] = C(h)$. Hence, process $\{Y(s): s \in R\}$ is *covariance stationary* if and only if (iff) the following two conditions hold for all $s,v \in R$,

(3.3.6) $$E[Y(s)] = \mu$$

(3.3.7) $$\| s-v \| = h \;\Rightarrow\; \text{cov}[Y(s),Y(v)] = C(h)$$

Note in particular from (3.3.7) that since $\text{var}[Y(s)] = \text{cov}[Y(s),Y(s)]$ by definition, and since $\| s-s \| = 0$, it follows that these random variables must also have a *common variance*, σ^2 given by

(3.3.8) $$\text{var}[Y(s)] = C(0) = \sigma^2 \quad , \quad s \in R$$

While these definitions are in terms of general spatial stochastic processes, $\{Y(s): s \in R\}$, our most important applications will be in terms of spatial residuals (random effects). With this in mind, notice that (3.3.6) together with (1.2.1) imply that every covariance stationary process can be written as

(3.3.9) $$Y(s) = \mu + \varepsilon(s)$$

so that each such process is associated with a unique *residual process*, $\{\varepsilon(s): s \in R\}$. Moreover, since $\text{cov}[Y(s),Y(v)] = \text{cov}[\varepsilon(s),\varepsilon(v)] = E[\varepsilon(s)\cdot\varepsilon(v)] - E[\varepsilon(s)]\cdot E[\varepsilon(v)]$, we see that $\{\varepsilon(s): s \in R\}$ must satisfy the following more specialized set of conditions for all $s,v \in R$:

(3.3.10) $$E[\varepsilon(s)] = 0$$

(3.3.11) $$\| s-v \| = h \;\Rightarrow\; E[\varepsilon(s)\varepsilon(v)] = C(h)$$

These are the appropriate covariance stationarity conditions for *residuals* that correspond to the stronger Homogeneity (3.3.2) and Isotropy (3.3.3) conditions in Section 3.3.1 above.

Note finally that even these assumptions are too strong in many contexts. For example (as mentioned above) it is often convenient to relax the isotropy condition implicit in (3.3.7) and (3.3.11) to allow directional variations in covariances. This can be done by requiring that covariances dependent only on the *difference* between locations, i.e., that for all $h = (h_1,h_2)$, $s-v = h \;\Rightarrow\; \text{cov}[Y(s),Y(v)] = C(h)$. This weaker stationarity condition is often called *intrinsic stationarity*. See for example [BG] (p.162), Cressie (1993, Sections

2.2.1 and 2.3) and Waller and Gotway (2004, p.273). However, we shall treat *only* the isotropic case [(3.3.7),(3.3.11)], and shall use these assumptions throughout.

3.3.3 Covariograms and Correlograms

Note that since the above covariance values, $C(h)$, are unique for each distance value, h, in region R, they define a function, C, of these distances which is designated as the *covariogram* for the given covariance stationary process.[16] But as with all random variables, the values of this covariogram are only meaningful with respect to the particular units in which the variables are measured. Moreover, unlike mean values, the values of the covariogram are actually in squared units, which are difficult to interpret in any case. Hence it is often more convenient to analyze dependencies between random variables in terms of (dimensionless) correlation coefficients. For any stationary process, $\{Y(s): s \in R\}$, the (product moment) *correlation* between any $Y(s)$ and $Y(v)$ with $\|s - v\| = h$ is given by the ratio:

$$(3.3.12) \qquad \rho[Y(s),Y(v)] = \frac{\text{cov}[Y(s),Y(v)]}{\sqrt{\text{var}[Y(s)]}\sqrt{\text{var}[Y(v)}} = \frac{C(h)}{\sqrt{C(0)}\sqrt{C(0)}} = \frac{C(h)}{C(0)}$$

which is simply a normalized version of the covariogram. Hence the correlations at every distance, h, for a covariance stationary process are summarized by a function called the *correlogram* for the process:

$$(3.3.13) \qquad \boxed{\rho(h) = \frac{C(h)}{C(0)} \quad , \quad s \in R}$$

Probably the most important application of correlograms is to allow comparisons between covariograms that happen to be in different units. One such application is illustrated in Section 7.3.5 below.

[16] To be more precise, if the set of all distances associated with pairs of locations in region R is denoted by $h(R) = \{h : \| s - v \| = h \text{ for some } s, v \in R\}$, then the covariogram, C, is a numerical function on $h(R)$. Note also that for the weaker form of *intrinsic stationarity* discussed above, the covariogram depends on the differences in both coordinates, $h = (h_1, h_2)$, and hence is a *two-dimensional* function in this case.