

X. Emery

On the consistency of the indirect lognormal correction

Abstract The indirect lognormal correction is a change-of-support model commonly used in geostatistical applications when dealing with additive variables, for which the upscaling amounts to arithmetic averaging. It was designed as a generalization of the lognormal correction that states the permanence of lognormality, but so far its internal consistency has not been proven in the general case. After a recall of the theoretical conditions that change-of-support models must honor, the concept of conventional income is introduced and used to establish the mathematical consistency of the indirect lognormal correction. However, the suitability of this model is questionable in many situations, in particular when the support effect is important or when the point-support distribution presents a zero effect, is not continuous or not positively skewed.

Keywords Change of support · Cartier's relation · Conventional income · Zero effect

1 Introduction

The support effect modeling is a key problem in several fields of applications, such as mining engineering, agricultural land management, forest inventories, pollution studies and image analysis: given a point-support distribution (histogram of the sampled values), one seeks the distribution of the values over a bigger support. Several models have been proposed in such a way, all of them based on a transformation of the sample distribution that accounts for the support effect, e.g. the affine correction, the mosaic correction and the discrete gaussian model (Matheron 1978, 1984c; Journel and Huijbregts 1978, p 471–475; Lantuéjoul 1990, p 54; Chilès and Delfiner

1999, p 431; Lajaunie 2000). This work aims at proving the consistency of a fourth model that is frequently used in geostatistical applications: the indirect lognormal correction (Isaaks and Srivastava 1989, p 472), which generalizes the so-called lognormal correction based on the hypothesis of permanence of lognormality.

2 Conditions for a consistent change-of-support modeling

Let us consider a stationary random field $\{Z_{\mathbf{x}}, \mathbf{x} \in R^d\}$ with positive values and define its average over a block support v with volume $|v|$:

$$Z_v = \frac{1}{|v|} \int_v Z_{\mathbf{x}} d\mathbf{x} \quad (1)$$

Equation (1) only applies to *additive variables*, for which the value of a block is defined as the arithmetic average of the point-support values within the block. This excludes variables such as solubility ratios (mining applications), pH (agronomy and soil sciences) or permeability (groundwater hydrology and petroleum engineering).

The distributions of both random variables $Z_{\mathbf{x}}$ and Z_v are linked together. Indeed, the mean is unchanged:

$$E(Z_v) = E(Z_{\mathbf{x}}) \quad (2)$$

whereas the variance of Z_v is smaller than the one of $Z_{\mathbf{x}}$ (because of Schwarz's inequality on the covariance function):

$$\text{var}(Z_v) = \frac{1}{|v|^2} \int_v \int_v \text{cov}\{Z_{\mathbf{x}}, Z_{\mathbf{x}'}\} d\mathbf{x} d\mathbf{x}' \leq \text{var}(Z_{\mathbf{x}}) \quad (3)$$

Now, despite a widespread belief, conditions (2) and (3) are not enough to guarantee the model consistency, as illustrated in Fig. 1: the proportion of extreme-high values (say, the values greater than 0.8) of the block-support distribution is not compatible with the point-support distribution. In brief, the point-support distribution induces additional constraints on the shape of the

X. Emery
Department of Mining Engineering,
University of Chile,
Avenida Tupper 2069, Santiago, Chile
E-mail: xemery@cec.uchile.cl
Tel.: +56-2-672-3504, +56-2-678-4498

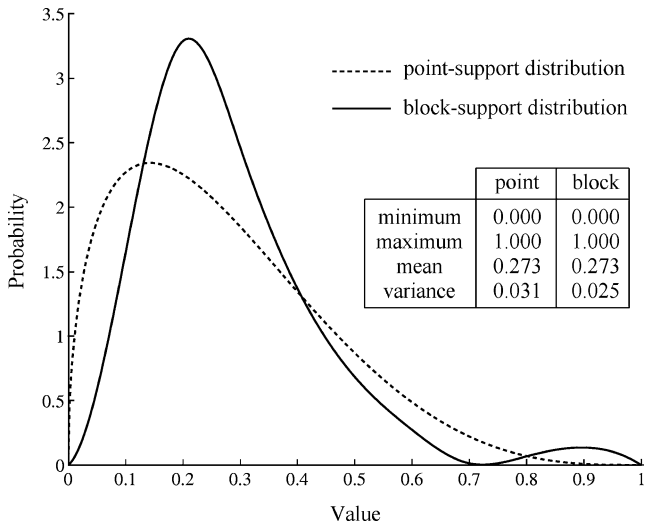
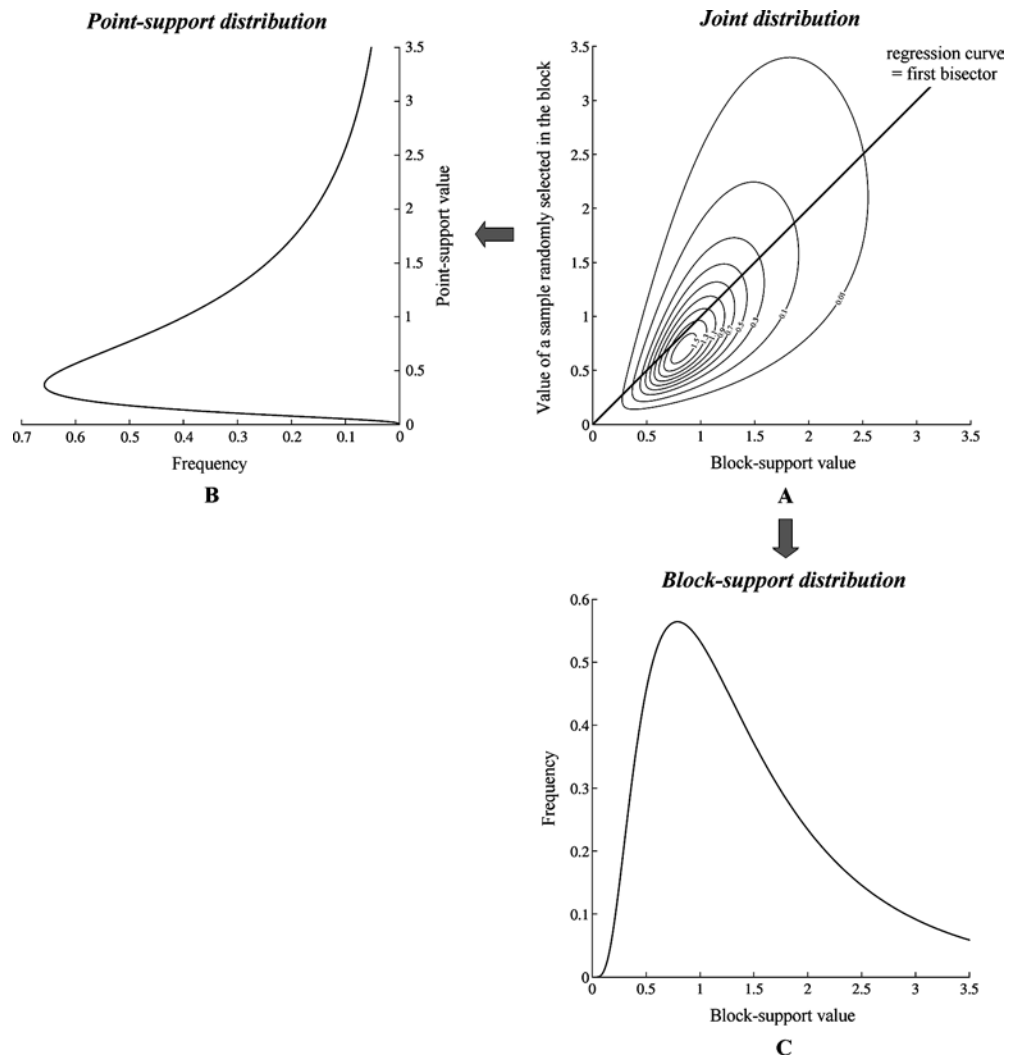


Fig. 1 An example of inconsistent change-of-support model

associated block distribution; for instance the extremal values of the latter must lie inside the range of the former.

To ensure the consistency of the change-of-support model, a third condition, known as *Cartier's relation*,

Fig. 2 A, Joint point-block distribution and B, C, marginal distributions at both supports



must be honored (Matheron 1984b; Lantuéjoul 1990, p 53; Chilès and Delfiner 1999, p 427): if \underline{x} stands for a random point uniformly distributed inside v , then

$$E(Z_{\underline{x}}|Z_v) = Z_v \tag{4}$$

Despite its apparent simplicity, Eq. (4) is quite restrictive, since it contains *per se* both relations (2) and (3) (Lantuéjoul, 1990 p 44–46) and entails several constraints on the shape and the extension of the block-support distribution with respect to the point-support one. Indeed, Cartier's relation states that, given a block with a known value, the value of a sample randomly selected in this block is expected to be equal to the block value, hence the regression curve of the joint point-block distribution coincides with the first bisector (Fig. 2A). Now, the marginal distributions are deduced from the joint distribution (Fig. 2B and 2C), so the constraint on the latter induces constraints on the formers. Cartier's relation is the basis of several change-of-support models, such as the discrete gaussian model (Matheron 1974, p 39; Chilès and Delfiner 1999, p 432) or the discrete isofactorial models based on nongaussian distributions (Matheron 1984a; Demange et al. 1987; Hu 1988).

3 The direct and indirect lognormal corrections

3.1 The lognormal correction

This model is based on an empirical observation that states the permanence of lognormality when the support increases (Matheron 1978, p 5; Journel and Huijbregts 1978, p 468). It consists in modeling the point-support histogram thanks to a lognormal distribution with mean m and variance σ_x^2 , then in assuming that the block-support variable has a lognormal distribution with the same mean but a smaller variance, so as to honor Eq. (2) and (3). Now, such a model is also consistent with Cartier's relation, since it matches the discrete gaussian model when the anamorphosis is an exponential function (Chilès and Delfiner 1999, p 433).

The lognormal correction can be summarized via the following relationship, which is an equality of the distributions (Isaaks and Srivastava 1989, p 473):

$$Z_v \stackrel{D}{=} aZ_x^b \tag{5}$$

with

$$b = [\ln(1 + \sigma_v^2/m^2) / \ln(1 + \sigma_x^2/m^2)]^{1/2}$$

$$a = m^{1-b} [1 + \sigma_v^2/m^2]^{-1/2} [1 + \sigma_x^2/m^2]^{b/2}.$$

The parameter b lies between 0 and 1 and is called *change-of-support coefficient*. The lower b , the stronger the support effect.

3.2 The indirect lognormal correction

The indirect lognormal correction generalizes the previous model by applying Eq. (5) even if the point-support distribution is not lognormal. In general, such a procedure is likely to alter the mean of the distribution, so that the corrected block-support distribution is rescaled to the point-support mean (Isaaks and Srivastava 1989, p 474):

$$Z_v \stackrel{D}{=} aZ_x^b \tag{6}$$

with $b = [\ln(1 + \sigma_v^2/m^2) / \ln(1 + \sigma_x^2/m^2)]^{1/2}$
 a calculated so that both distributions share the same mean (m).

Equation (6) provides an identity between the cumulative distribution functions F_x and F_v of the variables at both supports:

$$F_v(z) = \text{Prob}\{Z_v < z\} = \text{Prob}\{Z_x < (z/a)^{1/b}\}$$

$$= F_x((z/a)^{1/b}) \tag{7}$$

3.3 Choice of the parameters

The parameters a and b in Eq. (5) give the correct block-support variance (σ_v^2) only if the point-support distribution is perfectly lognormal. In the general case

[Eq. (6)], although the mean is unchanged by the support correction, the block-support distribution has a variance that differs from σ_v^2 (Isaaks and Srivastava 1989, p 487).

Henceforth, let us define the noncentered moments of the point-support and block-support distributions:

$$\forall \alpha > 0, \mu_x^{(\alpha)} = E(Z_x^\alpha) \text{ and } \mu_v^{(\alpha)} = E(Z_v^\alpha) \tag{8}$$

These moments are linked together [Eq. (7)]:

$$\mu_v^{(\alpha)} = \int_0^{+\infty} z^\alpha F_v(dz) = \int_0^{+\infty} (au^b)^\alpha F_x(du) = a^\alpha \mu_x^{(b\alpha)}$$

with $u = (z/a)^{1/b}$ (9)

In particular:

$$\begin{cases} \mu_v^{(1)} = m = a\mu_x^{(b)} \\ \mu_v^{(2)} = m^2 + \sigma_v^2 = a^2\mu_x^{(2b)} \end{cases} \tag{10}$$

This allows determining the value of the parameter b so as to honor a prescribed mean m and a block-support variance σ_v^2 :

$$1 + \frac{\sigma_v^2}{m^2} = \frac{\mu_x^{(2b)}}{(\mu_x^{(b)})^2} = \varphi(b) \tag{11}$$

For absolutely continuous distributions, Eq. (11) has at least one solution in $[0,1]$, since one has:

$$\varphi(0) = 1 \leq 1 + \frac{\sigma_v^2}{m^2} \leq 1 + \frac{\sigma_x^2}{m^2} = \varphi(1) \tag{12}$$

Actually, as will be seen in the next section, the solution of Eq. (11) is unique, which proves that the mapping $b \rightarrow \varphi(b)$ is increasing.

However, this conclusion does no longer hold if the point-support distribution is not continuous. For instance, consider a distribution with a proportion q of zeroes ("zero effect"); then

$$\lim_{\substack{b \rightarrow 0 \\ b > 0}} \varphi(b) = \frac{1}{1-q} > 1 \tag{13}$$

so that Eq. (11) has no solution if

$$q > \frac{\sigma_v^2}{m^2 + \sigma_v^2} \tag{14}$$

Application to the lognormal model

A point-support lognormal value Z_x with mean m and variance σ_x^2 can be written as follows:

$$Z_x = m \exp(sY_x - s^2/2) \tag{15}$$

where $s^2 = \ln\left(1 + \frac{\sigma_x^2}{m^2}\right)$ is the logarithmic variance (Journel and Huijbregts 1978, p 480).

Y_x is a standard gaussian random variable (zero mean and unit variance).

Hence, for any positive scalar ω , the moment of order ω of Z_x is

$$\mu_x^{(\omega)} = m^\omega \exp(-\omega s^2/2) E\{\exp(\omega s Y_x)\}$$

with $E\{\exp(\omega s Y_x)\} = \exp(\omega^2 s^2/2)$ (16)

so that one finally obtains

$$\mu_x^{(\omega)} = m^\omega \left(1 + \frac{\sigma_x^2}{m^2} \right)^{\frac{\omega(\omega-1)}{2}} \tag{17}$$

Solving Eq. (11) provides the coefficient b given in Eq. (5):

$$1 + \frac{\sigma_v^2}{m^2} = \left(1 + \frac{\sigma_x^2}{m^2} \right)^{b^2}, \quad \text{i.e. } b = \sqrt{\frac{\ln(1 + \sigma_v^2/m^2)}{\ln(1 + \sigma_x^2/m^2)}} \tag{18}$$

4 Consistency of the indirect lognormal correction

4.1 Mathematical point of view

The indirect lognormal correction accounts for the equality between the mean values [Eq. (2)] and the reduction of variance [Eq. (3)]. However, so far, the consistency of such a model has not been fully proven. More specifically, the question is: does the indirect lognormal correction honor Cartier's relation [Eq. (4)]?

A convenient way to answer this question is to use the concept of conventional income, which has been introduced in mining geostatistics to quantify the notion of *selectivity*. The conventional income at a threshold z is defined as the sum of the ore tonnage from z to infinity (Matheron 1984b, p 422; Lantuéjoul 1990, p 33; Chilès and Delfiner 1999, p 422):

$$\begin{aligned} B_x(z) &= \int_z^{+\infty} T_x(u) du \text{ with } T_x(u) \\ &= \text{Prob}\{Z_x \geq u\} = 1 - F_x(u) \end{aligned} \tag{19}$$

The conventional income $B_x(z)$ is a convex and nonincreasing function on R_+ that tends to zero at infinity and

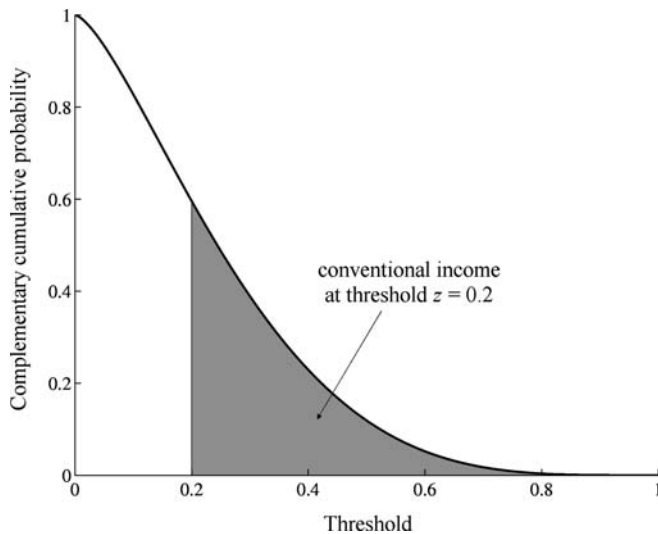


Fig. 3 Definition of the conventional income from the complementary cumulative distribution function

completely characterizes the distribution of Z_x . Similarly, for the block support, one can define:

$$\begin{aligned} B_v(z) &= \int_z^{+\infty} T_v(u) du \text{ with } T_v(u) \\ &= \text{Prob}\{Z_v \geq u\} = 1 - F_v(u) \end{aligned} \tag{20}$$

Since Z_x and Z_v have positive values, their common average is equal to the conventional incomes at threshold $z = 0$:

$$m = B_x(0) = B_v(0) \tag{21}$$

Now, Cartier's relation [Eq. (4)] is equivalent to the double condition:

- i) Z_x and Z_v have the same mean value;
- ii) the distribution of Z_v is less selective than the distribution of Z_x , which amounts to honor the following inequality between their conventional incomes (Matheron 1984b, p 424; Lantuéjoul 1990, p 44; Chilès and Delfiner 1999, p 425):

$$\forall z \in R_+, B_v(z) \leq B_x(z) \tag{22}$$

The parameter a in Eq. (6) is defined in order to ensure condition i). Then the indirect lognormal correction will constitute a consistent change-of-support model if and only if condition ii) is met. To check inequality (22), let us define the difference between the conventional incomes at both supports:

$$\forall z \in R_+, \Delta(z) = B_v(z) - B_x(z) \tag{23}$$

By differentiating this quantity, it comes:

$$\forall z \in R_+, \Delta'(z) = F_v(z) - F_x(z) = F_x((z/a)^{1/b}) - F_x(z) \tag{24}$$

Suppose now that F_x is a strictly increasing function. Then the derivative has the same sign as $(z/a)^{1/b} - z$.

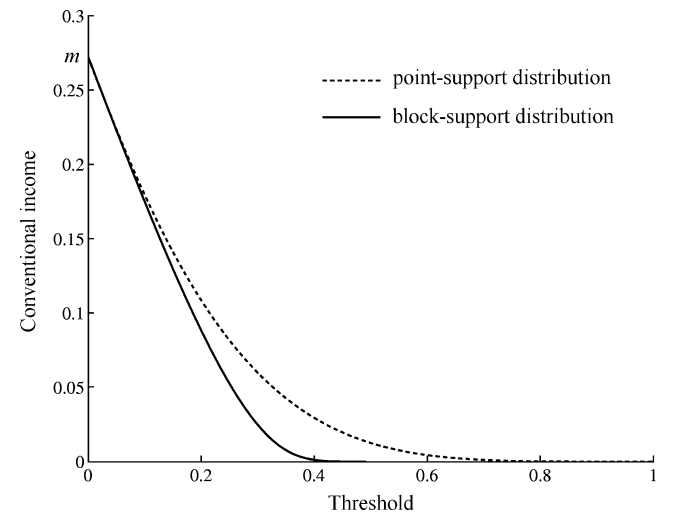


Fig. 4 Evolution of the conventional income with the support effect

Since b belongs to the interval $[0,1]$ [Eq. (12)], one obtains:

$$\forall z \in \mathbb{R}_+, \begin{cases} \Delta'(z) = 0 & \text{if } z = 0 \text{ or } z = a^{1/(1-b)} \\ \Delta'(z) < 0 & \text{if } z \in [0, a^{1/(1-b)}] \\ \Delta'(z) > 0 & \text{if } z \in [a^{1/(1-b)}, +\infty] \end{cases} \quad (25)$$

Hence, the maximal values of $\Delta(z)$ are $\Delta(0) = 0$ and $\Delta(+\infty) = 0$. This statement implies that $\Delta(z)$ is negative over the whole range of threshold values. The conclusion holds even if F_x is not a strictly increasing function, because it can always be written as the limit of a set of strictly increasing functions, so that inequality (22) remains asymptotically true.

In conclusion, the indirect lognormal correction honors Cartier's relation and constitutes a mathematically consistent change-of-support model for positive variables.

Incidentally, the concept of conventional income enables to prove that the solution of Eq. (11) is unique. Indeed, let us assume that two pairs of parameters $\{a, b\}$ and $\{a', b'\}$, with $b' < b$, are such that $Z_1 = aZ_x^b$ and $Z_2 = a'Z_x^{b'}$ have the same mean m and variance σ_v^2 . Since $b' < b$, Z_2 is itself deduced from Z_1 by an indirect lognormal correction, with a reduction factor equal to b'/b . Therefore, the conventional income of Z_2 is smaller than the one of Z_1 [Eq. (22)], so their variances are different, which contradicts the initial assumption (the variance is twice the area located below the conventional income curve and above the line $B(z) = m - z$) (Chilès and Delfiner 1999, p 421).

4.2 Practical issues

Despite its theoretical consistency, the indirect lognormal correction suffers from several practical limitations that are detailed hereafter.

- 1) The application of the support correction should be limited to small supports. Indeed, when b tends to zero (very large support), Eq. (6) becomes

$$Z_v \stackrel{D}{=} a[1 + b \ln(Z_x)] \quad (26)$$

Except for the case where Z_x has a lognormal distribution, the asymptotic distribution is not gaussian, as expected by the central limit theorem under strong mixing conditions for the random field $\{Z_x, \mathbf{x} \in \mathbb{R}^d\}$ (Gordin, 1969). This limitation is not proper to the indirect lognormal correction, since it is shared by the affine and mosaic corrections, but not by the discrete gaussian model (Chilès and Delfiner 1999, p 432–434).

- 2) The model cannot handle a “zero effect” in a suitable way. Indeed, let us assume that a proportion q of the point-support values is equal to zero:

$$F_x(0) = 0 \text{ and } \lim_{\substack{z \rightarrow 0 \\ z > 0}} F_x(z) = q \quad (27)$$

Then, according to Eq. (7), the block-support distribution has the same proportion of zeroes:

$$F_v(0) = 0 \text{ and } \lim_{\substack{z \rightarrow 0 \\ z > 0}} F_v(z) = q \quad (28)$$

Such a situation is unrealistic: one would expect the zero effect to decrease when the support increases. To overcome this difficulty, other change-of-support corrections should be preferred, like the mosaic or the nongaussian isofactorial models (Matheron 1984c; Demange et al. 1987; Chilès and Delfiner 1999, p 434).

- 3) The block-support distribution can be more skewed than the point-support distribution. To illustrate this statement, let us assume that the latter is a beta distribution and compare the skewness coefficients of both the point-support and the block corrected distributions (the choice of a beta distribution is motivated by the fact that it takes varied shapes according to the input parameters). The point-support skewness coefficient is defined by

$$\gamma_1^{(x)} = \frac{E[(Z_x - m)^3]}{\{E[(Z_x - m)^2]\}^{3/2}} = \frac{\mu_x^{(3)} - 3\mu_x^{(2)}\mu_x^{(1)} + 2[\mu_x^{(1)}]^3}{\{\mu_x^{(2)} - [\mu_x^{(1)}]^2\}^{3/2}} \quad (29)$$

Using Eq. (9), the block-support skewness coefficient is

$$\begin{aligned} \gamma_1^{(v)} &= \frac{\mu_v^{(3)} - 3\mu_v^{(2)}\mu_v^{(1)} + 2[\mu_v^{(1)}]^3}{\{\mu_v^{(2)} - [\mu_v^{(1)}]^2\}^{3/2}} \\ &= \frac{\mu_x^{(3b)} - 3\mu_x^{(2b)}\mu_x^{(b)} + 2[\mu_x^{(b)}]^3}{\{\mu_x^{(2b)} - [\mu_x^{(b)}]^2\}^{3/2}} \end{aligned} \quad (30)$$

For a beta distribution with parameters $\{\alpha, \beta\}$, the noncentered moment of order ω is given by the following formula (Papoulis 1984, p 147):

$$\mu_x^{(\omega)} = \frac{\Gamma(\alpha + \beta)\Gamma(\alpha + \omega)}{\Gamma(\alpha)\Gamma(\alpha + \beta + \omega)} \quad (31)$$

Equations (29) to (31) enable to express the ratio $\gamma_1^{(v)}/\gamma_1^{(x)}$ as a function of α, β and b . A ratio greater than one indicates that the block-support distribution is more skewed than the point-support one, whereas a negative ratio indicates an inversion of the skewness sign. Now, these unintuitive situations are quite frequent, as shown in Fig. 5. With negatively skewed point-support distributions, the block-support skewness is more negative (cases $\alpha = 10, \beta = 2$ and $\alpha = 1, \beta = 0.5$), whereas positively skewed point-support distributions tend to give negatively skewed block-support distributions when the support correction is important (cases $\alpha = 2, \beta = 10$ and $\alpha = 0.5, \beta = 1$). In all cases, the indirect lognormal correction decreases the skewness of the point-support distribution. In particular, when nonpositive, the skewness coefficient increases in absolute value: for instance, a perfectly symmetric distribution, like a uniform distribution in $[0, 1]$, would be transformed into a negatively skewed distribution.

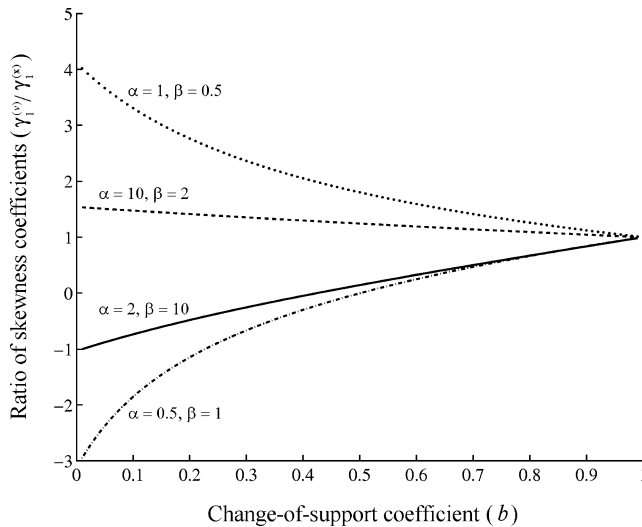


Fig. 5 Ratio of the block-support and point-support skewness coefficients as a function of the change-of-support coefficient

A consequence of these statements is that the indirect lognormal correction should not be applied in a local framework, for instance in association with indicator kriging. Indeed, the local distributions (i.e. the distributions conditional to a set of neighboring data) generally have extremely varied shapes and their skewness can strongly differ from the one of the global distribution, hence the risk of an unrealistic support correction increases. Actually, performing a change of support with indicator kriging is a complex (and still unsolved) problem, because of the need of an appropriate support correction for the local distributions. In general, the affine and mosaic corrections are not realistic options either.

The last two statements are limitations of lesser importance, so they will be mentioned without insisting on them.

- 4) The shape of the corrected distribution is not invariant when the point-support distribution is shifted. Moreover, the minimal value is arbitrarily increased by the correction [Eq. (6)], excepted when this minimum is equal to zero. For instance, a shifted lognormal distribution no longer provides a lognormal distribution for the block-support values, and the minimal values of both distributions are different (the amplitude of the difference even depends on the shift parameter that is applied). Consequently, in practice, if the sample distribution clearly begins apart from zero, it should be shifted before performing the change of support.
- 5) The model is not suitable for discrete variables, unless the number of classes is large and can be seen as a discretization of a continuous distribution. For instance, if Z_x only takes a small number of values, so does Z_b [Eq. (6)], which is clearly improbable. In such a situation, one may use a mixed change-of-support model, where the point-support variable is discrete

while the regularized variable is continuous. Examples are the negative binomial/gamma model (Matheron 1984a, p 99) and the discrete Jacobi/beta model (Matheron 1984d, p 58). An alternative model is the uniform mixture introduced by Matheron (1985, p 164).

To summarize, the indirect lognormal correction is sound when the point-support distribution departs from the lognormal one while keeping its main features, that is: continuous and positively skewed variables with a range of variation from zero to a maximum (and hence no zero effect). And like the lognormal correction, its validity is limited to a moderate change of support.

5 Conclusions

The indirect lognormal correction is a mathematically consistent change-of-support model since it honors Cartier's relation. However, the user should beware of the misuses of this model, in particular when dealing with noncontinuous variables, zero effects or with negatively skewed variables. In practice, to avoid an unrealistic correction, its application is restricted to small supports and to variables whose histogram is "close" to a lognormal distribution. In other cases, alternative models should be preferred, such as the *discrete gaussian model* that also generalizes the direct lognormal correction but has proved to be suitable in a wide variety of situations (Maréchal 1975; Matheron 1985, p 153; Kavourinos 1987; Chilès and Delfiner 1999, p 447; Lajaunie 2000).

References

- Chilès JP, Delfiner P (1999) Geostatistics: Modeling spatial uncertainty. Wiley, New York, pp 696
- Demange C, Lajaunie C, Lantuéjoul C, Rivoirard J (1987) Global recoverable reserves: testing various changes of support models on uranium data. In: Matheron G, Armstrong M (eds) Geostatistical Case Studies. Reidel, Dordrecht
- Gordin MI (1969) The central limit theorem for stationary processes. Soviet Mathematics Doklady 10:1174–1176
- Hu LY (1988) Mise en œuvre du modèle gamma pour l'estimation de distributions spatiales, Doctoral thesis, Centre de Géostatistique, Ecole des Mines de Paris, Fontainebleau
- Isaaks EH, Srivastava RM (1989) An introduction to applied geostatistics. Oxford University Press, New York, pp 561
- Journel AG, Huijbregts CJ (1978) Mining geostatistics. Academic Press, London, pp 600
- Kavourinos C (1987) The grade-tonnage curves for a zinc mine in France. In: Armstrong M, Matheron G (eds) Geostatistical Case Studies. Reidel, Dordrecht
- Lajaunie C (2000) Introduction to the modelling of the support effect. In: Pandalai HS, Saraswati PK (eds) Geological data analysis: statistical methods. Hindustan Publishing, New Delhi
- Lantuéjoul C (1990) Cours de sélectivité, course notes C-140, Centre de Géostatistique, Ecole des Mines de Paris, Fontainebleau, pp 72
- Maréchal A (1975) Forecasting a grade-tonnage distribution for various panel sizes. In: 13th International Symposium on Computer Applications in the Mineral Industries (APCOM), Clausthal, West Germany

- Matheron G (1974) Les fonctions de transfert des petits panneaux, internal report N-395, Centre de Géostatistique, Ecole des Mines de Paris, Fontainebleau, pp 73
- Matheron G (1978) L'estimation globale des réserves récupérables, course notes C-75, Centre de Géostatistique, Ecole des Mines de Paris, Fontainebleau, pp 28
- Matheron G (1984a) Modèles isofactoriels et changement de support. *Sciences de la Terre Série Informatique Géologique* 18:71–123
- Matheron G (1984b) The selectivity of the distributions and the “second principle of geostatistics”. In: Verly G, David M, Journel AG, Maréchal A (eds) *Geostatistics for Natural Resources Characterization*. Reidel, Dordrecht
- Matheron G (1984c) Changement de support en modèle mosaïque. *Sciences de la Terre Série Informatique Géologique* 20:435–454
- Matheron G (1984d) Une méthodologie générale pour les modèles isofactoriels discrets. *Sciences de la Terre Série Informatique Géologique* 21:1–64
- Matheron G (1985) Change of support for diffusion-type random functions. *Math Geol* 17(2):137–165
- Papoulis A (1984) *Probability, random variables and stochastic processes*. McGraw-Hill Book Co., New York, pp 583