Guaranteeing Proportions with Indicator Kriging

John G. Manchuk and Clayton V. Deutsch

Order relations are a common problem with indicator methods in geostatistics. Estimated proportions are often negative and rarely sum to unity. Various methods exist to adjust results to meet these constraints; however, the objective of such methods is not to minimize the estimation variance. It is possible with optimization methods to estimate a vector of proportions that are all positive, sums to unity, and has the property of minimum estimation variance. Lagrange multipliers are used to combine constraints into the variance equation and a combinatorial technique is used to determine which multipliers are needed to obtain feasible solutions. Despite the increased complexity, execution time is not unreasonable for a limited number of categories.

Introduction

There is no intrinsic property of indicator kriging that constrains estimated proportions to be positive and have unit sum. Occurrences where indicator kriging actually produces proportions meeting these constraints are purely random and highly unlikely. However, the deviation from these constraints is typically not substantial so quick fixes are not immediately rejected – negative proportions are typically set to zero and those remaining are standardized. Fixes of this nature are not necessarily optimal. As with any form of kriging, the underlying objective when dealing with indicators is to minimize the error variance. This objective should be maintained while searching for the constrained solution to indicator kriging.

An identical problem to this has been encountered with compositional data, where a set of variables that are all positive sum to a whole. Each variable is a percentage or proportion of the whole composition. The method of compositional kriging was developed, initially by De Gruijter, Walvoort, and van Gaans (1997) and reiterated by Walvoort and De Gruijter (2001). Compositional kriging can thus be applied to indicator kriging where the estimates for each indicator class correspond to probabilities that must be positive and sum to unity.

In this paper, the methods developed for compositional kriging are applied to categorical data. Constraints are developed for simple indicator kriging and the resulting system is solved with quadratic programming (Papadimitriou and Steiglitz, 1998; Boyd and Vandenberghe, 2004). The algorithm has been embedded within a version of sequential indicator simulation (Deutsch and Journel, 1998) for analysis.

Background

Any form of kriging forms a quadratic program (QP). The objective is to minimize a function with a quadratic term, i.e. the estimation variance, under the constraint of a linear system of equations. This minimization problem is given by (1) where σ_E^2 and σ^2 are the estimation variance and the global variance respective, C_{ij} is a matrix of covariance values between two samples at locations x_i and x_j, λ_i is the solution vector or kriging weights, and c_i is a vector of covariance values between the sample at x_i and the location of interest, x. $\lambda^T C\lambda$ is the quadratic term.

$$\begin{aligned} \min \quad & \sigma_E^2 = \sigma^2 + \sum_i \sum_j \lambda_i \lambda_j C_{ij} - 2\sum_i \lambda_i c_i \\ & \quad = \sigma^2 + \lambda^T C\lambda - 2\lambda^T c \\ \text{s.t.} \quad & C\lambda = c \end{aligned} \tag{1}$$

This is a special case of QP since its solution is determined by solving the system $C\lambda = c$. However, for indicators the solution to (1) is not necessarily valid. Consider simple indicator kriging with M categories and N sample data. The estimate, $\phi_k^*(x)$, is a probability that the value at x belongs to class k (2), where $I(x_i,k)$ is the membership of data x_i to class k and ϕ_k is the global proportion of class k. All k=1,…,M estimates must therefore be positive and sum to unity.

$$\phi_k^* = \sum_{i=1}^{N} I(x_i,k)\lambda_{ik} + \left(1 - \sum_{i=1}^{N} \lambda_{ik}\right)\phi_k, k = 1,...,M \tag{2}$$

A typical implementation of indicator kriging would solve (1) for each class k. M systems of equations of the form (3) would be built and solved, the resulting solution being substituted into (2) giving a vector of estimates ϕ^*. There is no intrinsic property of this method that ensures the sum of ϕ^* is unity and that each ϕ_k^* is positive and between zero and one. The QP must be reformulated with these constraints in mind.

$$C_k \lambda_k = c_k, k = 1,...,M \tag{3}$$

For simple indicator kriging, the constrained QP must be reformulated with the constant sum constraint (4) and positivity constraint (5). A constraint to ensure ϕ_k^* are less than one is not required since the (4) and (5) together ensure this.

$$f_0 = \sum_{k=1}^{M} \phi_k^* = \sum_{k=1}^{M} I(x,k)^T \lambda_k + \left(1 - \sum \lambda_k\right)\phi_k = 1 \tag{4}$$

$$f_{\mathrm{k}} = \phi_k^* = \sum_{i=1}^{N} I(x_i,k)\lambda_{ik} + \left(1 - \sum_{i=1}^{N} \lambda_{ik}\right)\phi_k \geq 0, k = 1,...,M \tag{5}$$

These constraints may be simplified into (6) and (7).

$$f_0 = \sum_{k=1}^{M} \left(I(x,k) - \phi_k\right)^T \lambda_k = 0 \tag{6}$$

$$f_k = \left(I(x,k) - \phi_k\right)^T \lambda_k + \phi_k \geq 0, k = 1,...,M \tag{7}$$

The optimal solution to (1) with these additional constraints can be solved using Lagrangian duality, where the objective function is reformulated with Lagrange multipliers associated with each constraint. If ψ and ξ_k, k=1,…,M are the multipliers associated with each constraint, the minimization problem becomes (8). Differentiating in terms of λ_k gives (9) and in terms of each of the multipliers recovers the constraint functions f_0 and f_k, which are minimized when f_0=0 and when $\left(I(x,k) - \phi_k\right)^T \lambda_k = -\phi_k$ for f_k.

$$\min \quad \sigma_E^2 = \sigma^2 + \lambda_k^T C_k \lambda_k - 2\lambda_k^T c_k + 2\psi f_0 + 2\xi f_k, k = 1,...,M \tag{8}$$

$$\frac{\partial \sigma_E^2}{\partial \lambda_k} = C_k \lambda_k - c_k + \psi(I(x,k) - \phi_k) + \xi_k(I(x,k) - \phi_k) + \xi_k \phi_k \tag{9}$$

Solving (8) is done by setting (9) to zero and the result can be formulated as a system of equations for each category k. Because ψ is in all k systems, they must be merged into a single system of the form (10) with k blocks, where s_k is the $N\times1$ vector $I(x,k)$-ϕ_k, and each C_k is an $N\times N$ block. This system cannot be directly solved as is. The inequality constraints associated with ξ_k may be active or inactive: If the solution vector $[\lambda_1,\ldots,\lambda_M,\psi]^T$ is feasible, such that all equality constraints are met, and $f_k = 0$, then constraint f_k is active. If

$f_k > 0$, it is inactive. So the system in (10) only requires an inequality constraint if ϕ_k^* is negative. Solving (10) therefore involves determining which set of active inequality constraints solves (8), where an empty set is valid. This is an NP-complete optimization problem, that is, a Non-deterministic Polynomial-time problem for which no polynomial time algorithms can solve.

$$\begin{bmatrix} C_1 & 0 & \cdots & 0 & s_1 & s_1 & 0 & \cdots & 0 \\ 0 & C_2 & \ddots & \vdots & s_2 & 0 & s_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 & \vdots & \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & C_M & s_M & 0 & \cdots & 0 & s_M \\ s_1^T & s_2^T & \cdots & s_M^T & 0 & 0 & \cdots & \cdots & 0 \\ s_1^T & 0 & \cdots & 0 & 0 & \ddots & \ddots & & \vdots \\ 0 & s_2^T & \ddots & \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 & \vdots & & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & s_M^T & 0 & \cdots & \cdots & 0 & 0 \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_M \\ \psi \\ \xi_1 \\ \xi_2 \\ \vdots \\ \xi_M \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_M \\ 0 \\ -\phi_1 \\ -\phi_2 \\ \vdots \\ -\phi_M \end{bmatrix} \tag{10}$$

A simplified version of (10) will be referred to in the remaining paper and is given by (11).

$$\begin{bmatrix} \mathbf{C} & \mathbf{s} & \mathbf{S} \\ \mathbf{s}^T & 0 & 0 \\ \mathbf{S}^T & 0 & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{\lambda} \\ \psi \\ \boldsymbol{\xi} \end{bmatrix} = \begin{bmatrix} \mathbf{c} \\ 0 \\ -\boldsymbol{\Phi} \end{bmatrix} \tag{11}$$

Methodology

The solution to (10) is a combinatorial optimization problem. First, the system is solved assuming all f_k are inactive. If the resulting $\phi_k^* \geq 0$, the solution is optimal and estimates sum to unity. However, if any f_k are violated then the optimal combination of active inequality constraints that minimizes (8) must be discovered. Like De Gruijter, Walvoort, and van Gaans (1997) and Walvoort and De Gruijter (2001), the method of Theil and Van de Panne is used. Each combination of constraints that must be activated is iteratively built into (10). That with the minimum estimation variance is the optimal combination.

Pseudo-code for the algorithm is as follows:

1. Solve $\begin{bmatrix} \mathbf{C} & \mathbf{s} \\ \mathbf{s}^T & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{\lambda} \\ \psi \end{bmatrix} = \begin{bmatrix} \mathbf{c} \\ 0 \end{bmatrix}$
2. Initialize $j = 0$; Calculate σ_0^2 using (8)
3. Calculate f_k^0 using (7), k=1,…,M
4. Initialize list L = combinatorial (all $f_k^0 < 0$) and build the set V_0: $f_k^0 < 0$
5. If $V_0 \neq \varnothing$, reset σ_0^2 large [$\varnothing$ denotes the empty set]
6. While $L(j)$ not empty
    a. Increment j
    b. Build **S** using active constraints in $L(j)$
    c. Solve $\begin{bmatrix} \mathbf{C} & \mathbf{s} & \mathbf{S}_j \\ \mathbf{s}^T & 0 & 0 \\ \mathbf{S}_j^T & 0 & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{\lambda} \\ \psi \\ \boldsymbol{\xi} \end{bmatrix}_j = \begin{bmatrix} \mathbf{c} \\ 0 \\ -\boldsymbol{\Phi}_j \end{bmatrix}$
    d. Calculate f_k^j, k=1,…,M and build set V_j: $f_k^j < 0$
    e. If $V_j \neq \varnothing \text{ and } V_j \subseteq V_{j-1}$ ignore combination $L(j)$ [$\subseteq$ denotes a subset of or equal to]

f. If $V_j \neq \varnothing$ and any $V_j \notin V_{j-1}$, update V_{j-1} with new active constraints and add new combinations to L.

g. If $V_j = \varnothing$, Calculate σ_j^2; if $\sigma_j^2 < \sigma_0^2$, save $\begin{bmatrix}\boldsymbol{\lambda} & \psi & \boldsymbol{\xi}\end{bmatrix}_*^T = \begin{bmatrix}\boldsymbol{\lambda} & \psi & \boldsymbol{\xi}\end{bmatrix}_j^T$ and $\sigma_0^2 = \sigma_j^2$

7. Return the optimal feasible solution $\begin{bmatrix}\boldsymbol{\lambda} & \psi & \boldsymbol{\xi}\end{bmatrix}_*^T$

Implementation

An existing version of sequential indicator simulation (sisim) from Deutsch and Journel (1998) was used as a host to the algorithm discussed in the previous sections. A Gaussian elimination solver that takes advantage of the sparse structure of (10) was written for step 1. For solving the system in step 6.c, the upper triangulated left hand side from solving step 1 is retained since it is constant. Only $\mathbf{S}_j$ is updated while cycling through L and blocks $\mathbf{S}_j$, $\mathbf{S}_j^T$, and $-\boldsymbol{\Phi}$ are incrementally reduced. The sparse structure only exists under the assumption that cross variograms between categories are not involved.

For the following analyses, the two dimensional data set from the Spatial Interpolation Comparison 97 was used for data coordinates. An arbitrary categorical variable with 5 categories was generated and arbitrary variograms assigned to each. Data will be plotted along with realizations from each version of sisim.

A run time comparison between the original version of sisim (SIS_0) and the version containing the new solution method (SIS_1) was conducted. The number of data involved in each system was increased from 2 to 32 while keeping the grid size and number of realizations to generate constant at 10,000 blocks and 10 realizations. It is intuitive that time will increase linearly with increasing grid size while the number of data involved is held constant. The time complexity with number of categories was not explored, but since the combinatorial of active constraints grows as 2^n, it is reasonable to assume run time would grow to impractical proportions with the number of categories. Results are as expected with the new solver being slower (Table 1). By using efficient solvers, both methods are approximately an $O(n^2)$ process; however, SIS_1 takes roughly 6.6 times longer. This can be seen in Figure 1.

Table 1: Time comparison

Number of data (n)	Time (s), SIS_0	Time (s), SIS_1
2	1	3
4	1	10
8	4	25
16	12	77
32	54	354

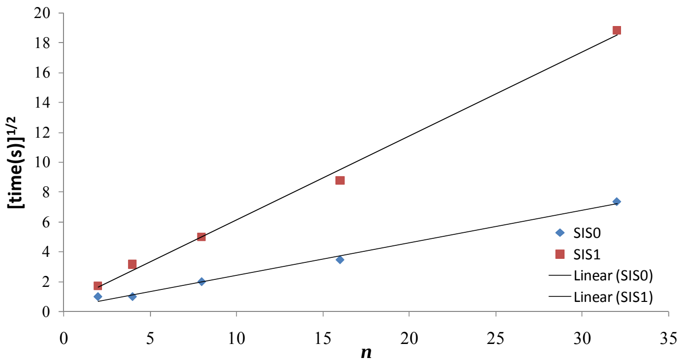

Figure 1: Execution time curves

Reproduction of the input proportions is also checked. Fifty realizations were generated over a 100×100 cell grid using 20 data for each kriging system. A plot of realization 10 for each method is shown in Figure 2, overlaid with the data. Global proportions from each realization were extracted to compare their statistics (Table 2). Both methods show acceptable reproduction of the global proportions over the fifty realizations with no error exceeding 2.7%. Inconsistent results are obtained when comparing the standard deviation of mean proportions for the two methods. Some cases show SIS_0 giving a higher standard deviation than SIS_1 and vise-versa. The interpretation is that the correction scheme used in SIS_0 to obtain positive probabilities and unit sum may result in proportions that would only be achievable by over or underestimating the actual kriging variance.

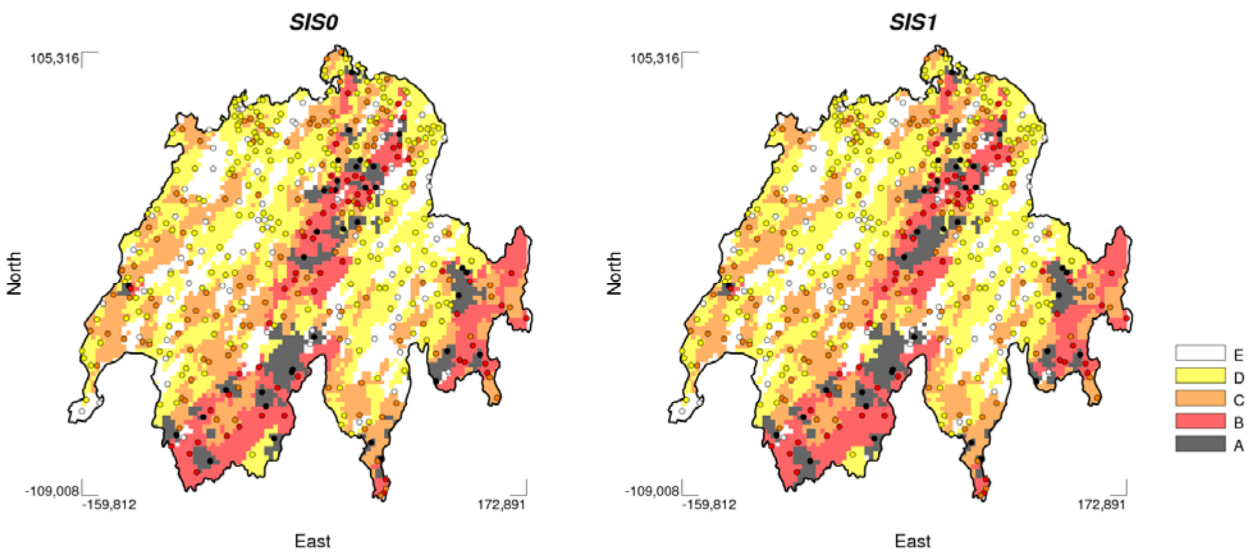

Figure 2: Realization 10 from each version of sisim

Table 2: Statistics of proportions for each version of sisim

	Category	Target	Min	Max	Mean	Standard Deviation
	A	0.07526	0.0504	0.1475	0.08005	0.019803
	B	0.12886	0.1028	0.1956	0.13510	0.022240
SIS_0	C	0.24666	0.1733	0.3197	0.24769	0.028547
	D	0.38826	0.2787	0.4180	0.36316	0.030676
	E	0.16096	0.1181	0.2632	0.17399	0.028848
	A	0.07526	0.0531	0.1665	0.08355	0.021296
	B	0.12886	0.1019	0.1940	0.13628	0.021666
SIS_1	C	0.24666	0.1632	0.2999	0.24430	0.027520
	D	0.38826	0.2689	0.4194	0.36147	0.031806
	E	0.16096	0.1217	0.2682	0.17439	0.027702

Conclusion

A constrained optimization technique, originally developed for compositional data, has been applied to indicator kriging. Constraints to maintain positive probabilities and to ensure estimates sum to unity can be incorporated into the systems of equations with Lagrange multipliers. A combinatorial technique is used to determine which constraints must be active to obtain a feasible solution. Results maintain the premise of minimum estimation variance, which is not necessarily maintained for other corrections involving order relation problems.

References

Walvoort, D.J.J. and De Gruijter, J.J. (2001) Compositional Kriging: a spatial interpolation method for compositional data: Mathematical Geology, Vol. 33, No. 8, 951-966

De Gruijter, J.J., Walvoort, D.J.J., and van Gaans, P.F.M., (1997) Continuous soil maps – a fuzzy set approach to bridge the gap between aggregation levels of process and distribution models: Geoderma, No. 77, 169-195

Boyd S., and Vandenberghe L. (2004) Convex optimization: Cambridge University Press, 716

Papadimitriou, C.H. and Steiglitz, K., (1998) Combinatorial optimization: algorithms and complexity: Dover, 512

Deutsch, C.V. and Journal, A.G., (1998) GSLIB: geostatistical software library and user's guide: Oxford University Press, 384