システム・ソフトウェア開発

対の一部を使うヴァリオグラムの計算における精度の低下と時間の短縮

正路 徹也*

Reduction of Accuracy and Time in Calculation of Variogram Using a Part of Pairs

Tetsuya SHOJI*

Abstract : It is essentially necessary to calculate square of the difference for every pair of given data in order to obtain a variogram, which is one of the most important functions in geostatistics. If data number is N , the number of pairs is $N(N-1)/2$. Accordingly, the calculation time increases proportionally to square of N . As general statistical data processing, it is possible to calculate a variogram using a part of data randomly collected from all data (say partial calculation using a part of data). In addition to this, variograms have been calculated using all data of a randomly rearranged sequence but using a part of their pairs of the following categories: 1) every datum pairs only limited data leading the sequence (partial calculation giving leading data priority), 2) every datum pairs only data neighboring in the sequence (partial calculation giving neighboring data priority), and 3) all data are divided into several groups of a similar size, and every datum pairs only the data belonging to the same group (partial calculation grouping data). When the similarity between two variograms are evaluated by correlation coefficient between them, variograms obtained by the partial calculations giving neighbors priority and using grouped data have high similarity to the variogram obtained by the complete calculation in the present data set, which is residues from a linear regression plane of submarine topography near Bonin islands ($N = 8891$). The correlation coefficient is higher than 0.99, if the partial calculation giving neighbor priority is carried out where the pair number is 1/100 of that of the complete calculation. If a correlation coefficient between two variograms whose neighbor numbers are different by 1 in the partial calculation is 0.9_s (s denotes number of 9 below the decimal point), it is expected that the correlation coefficient between the variograms in the partial and complete calculations is higher than 0.9_{s/2}.

Keywords: Variogram, Calculation, Partial, Time, Elevation, Accuracy, Efficiency

1. はじめに

地球科学データの空間的連続性の解析に用いられるヴァリオグラムは、次の式(1)で示される $\gamma(\mathbf{h})$ のように、2点を結ぶ距離ベクトル $\mathbf{h}$ の関数として、測定値から取り出された2個のデータからなる対について、それらの差の二乗を計算することによって得られる。

$$\gamma(\mathbf{h}) = \frac{\sum \{w(\mathbf{x} + \mathbf{h}) - w(\mathbf{x})\}^2}{p(\mathbf{h})} \quad (1)$$

ここで、 $w(\mathbf{x})$ と $w(\mathbf{x} + \mathbf{h})$ はそれぞれ地点ベクトルが $\mathbf{x}$ と $(\mathbf{x} + \mathbf{h})$ の地点における測定値、 $p(\mathbf{h})$ は距離が $\mathbf{h}$ である2地点の対の数である。したがって、計算に使われる対の数の総和 $p^\Sigma = \sum_{\mathbf{h}} p(\mathbf{h})$ は、データ数が N の場合、 $N(N-1)/2$ 個

である(第1図a右上の三角形の打点部)[注1]。このことは、ヴァリオグラムの計算時間がほぼデータ数の2乗に比例して長くなることを意味する。このため、データが多量に存在する場合、何らかの簡易的計算法を適用して、計算時間を短縮することが望まれる。ここでは、 $N(N-1)/2$ 個ある対の一部を使うヴァリオグラムの部分計算(partial calculation)を試み、結果をすべての対を使う完全計算(complete calculation)の結果と比較して、その精度を評価した。

部分計算の精度評価に用いたデータは、東京大学白鳳丸のKH-03-3航海で取得された小笠原母島付近の北緯21.21~21.24°(約3.34km)、東経142.87~142.90°(約3.00km)の範囲にある8891点の標高である(ほぼ南北に伸びる走査

第1図. 実験ヴァリオグラムの計算で使われるデータ対を示す模式図. 打点部が計算で使われる対を示し, 図aの右上が完全計算, 図bの左下と右上がそれぞれ一部データ部分計算と先頭優先部分計算, 図cの左下と右上がそれぞれ近隣順部分計算とグループ化部分計算の場合である.

線の間隔は約0.024km, 走査線上の測点の間隔は約0.048km). ただし, この範囲では南西から北東へ水深が深くなる傾向があるので, この傾向を平面で最小二乗近似し, それからの残差 (すなわち回帰残差) をヴァリオグラム計算のデータとした. ヴァリオグラムの計算条件は, 距離間隔が0.1km, 許容範囲が0.2kmの円である.

以下で紹介するヴァリオグラムはすべてMS-Excel/VBA (正路, 2002) で動くマクロプログラムで計算した (計算機はDell OptiPlex SX270). ところで, ヴァリオグラムは, 式(1)のように距離ベクトル $\mathbf{h}$ の関数として与えられる. このベクトル $\mathbf{h}$ を極座標 (h, θ) で表し, θ を固定してヴァリオグラムを動径 h (距離) の関数 $\gamma(h, \theta = \text{const})$ として与えた場合を方向ヴァリオグラムという. また, θ を無視してヴァリオグラムを距離 h のみの関数 $\gamma(h)$ として与えた場合を全方向ヴァリオグラムという (ここでは両者を合わせて極ヴァリオグラムと呼ぶ). 一方, ベクトル $\mathbf{h}$ を直交座標 (ξ, η) で表し, ヴァリオグラムを各成分の関数 $\gamma(\xi, \eta)$ で表した場合を, ここでは直交ヴァリオグラムと呼ぶ. 部分計算の評価は, このうち全方向実験ヴァリオグラムと直交実験ヴァリオグラムで行った. また, 方向ヴァリオグラムに球モデルを当てはめて求められたレンジも評価の対象とした.

2. 部分計算の方法

実験ヴァリオグラムを計算するためのデータは, 取得時の条件を反映して, 通常何らかの規則に従って並んでいる. これに対し, 対の一部を使う部分計算では, この規則が現れないようにデータをランダムに並べ替えておく必要がある. そこで, 以下の記述は, データが既にランダムに並べ替えられていることを前提に進める. このため, 「最初から n 番目までのデータ」は「ランダムに選ばれた n 個のデータ」と同義である.

部分計算の1つの方法としては, 通常の統計処理のランダムサンプリングと同様に, 最初から n 個 ($n \ll N$) のデータを計算に用いることが考えられる (第1図b左下の三

角形内の打点部). 以下, このデータの一部を使う計算に対しては, 「一部データ部分計算」と呼ぶ. この一部データ部分計算で用いられる対の数 P_n は次式で与えられる.

$$P_n = n(n-1)/2 \approx P(n/N)^2 \quad (2)$$

ここで, P は完全計算における対の数である. 一部データ部分計算は, N 個あるデータのうち n 個のデータしか使っておらず, 折角存在する $(N-n)$ 個のデータが無視される.

多くのデータが無視されるという一部データ部分計算の欠点を回避するために, データはすべて使うが, 総数で $N(N-1)/2$ ある対についてその一部しか使わない部分計算も試みた. 対をなすデータの計算では, 通常2重のループが使われる. 単純には, 外のループのカウンターを i , 内のループのカウンターを j で表して, i を1から $(N-1)$ まで, j を $(i+1)$ から N まで変化させると, ヴァリオグラムの完全計算が実行できる. このとき, 外のループのカウンターの i を n で止めると, 第1図b右上の三角形内の打点部で示されるように, データはすべて使うが, 対は一部しか使わない部分計算が可能となる. この計算方法は, データ並びの先頭から n 番目までのデータを優先しているので, 「先頭優先部分計算」と呼ぶ. この先頭優先部分計算で用いられる対の数 P_n は次式で与えられる.

$$P_n = nN - n(n+1)/2 \approx nN \quad (n \ll N) \approx 2P(n/N) \quad (3)$$

先頭優先部分計算では, n 番目までのデータは他のすべてのデータと対を組むのに対し, $(n+1)$ 番目からのデータは n 番目までのデータとしか対を組まないのので, データによって対を組む回数が極端に異なるという欠点がある.

データによって対を組む回数が極端に異なるという先頭優先部分計算の不都合を解消するためには, 各データについてそこから n 個離れた位置のデータまでの対について計算を実行する方法が考えられる. この方法だと, 第1図c左下の三角形内の打点部のうち, 対角位に近い部分の対が計算に用いられる. このため, 1番目と N 番目という両端

第2図. 実験ヴァリオグラムの部分計算で使われた対 (打点部) を示す模式図. 左から一部データ部分計算, 先頭優先部分計算, 近隣順序部分計算, グループ化部分計算. 各図とも, 左下の三角形が計算に使われた対を, 右上の三角形が元の並びでの対を示す. 一部データ部分計算と先頭優先部分計算では明らかに対はランダムには採用されていない. これに対し, 近隣順序部分計算とグループ化部分計算では対がランダムに採用されているように見えるが, 各データとも近隣順序部分計算では4回, グループ化部分計算では3ないし4回使われており, 計算に使う対の採用はランダムではない.

に近い位置のデータは, 計算に使われる回数が少なくなる. しかし, N 番目のデータの隣に1番目のデータがあると考え, 第1図c左下の三角形内の左下端付近の打点部で表される対も計算に使われるので, すべてのデータが $2n$ 回利用される. この計算方法を, 以下「近隣順序部分計算」と呼ぶ. この近隣順序部分計算で用いられる対の数 P_n は次式で与えられる.

$$P_n = nN \quad (n < N/2) \quad (4)$$

$$\approx 2P(n/N)$$

データはすべて使うが, 対は一部しか使わないという部分計算として, 第1図c右上の三角形内の打点部で示してあるデータを m 組に分ける計算も試みた. 以下, この計算を「グループ化部分計算」と呼ぶ. グループ化部分計算は次の手順で行う. ①データを m 個のグループに分ける. このとき, N を m で割った商を q_m 余りを r_m として, グループ1からグループ r_m まではデータの先頭から順次 $(q_m + 1)$ 個ずつとり, グループ $(r_m + 1)$ からグループ m までは q_m 個ずつとる. これにより, グループ1からグループ r_m までは $(q_m + 1)$ 個, グループ $(r_m + 1)$ からグループ m までは q_m 個のデータで構成される. ②グループ k のヴァリオグラム $\gamma_k(\mathbf{h})$ を次式で計算する.

$$\gamma_k(\mathbf{h}) = \sum \{w(\mathbf{x} + \mathbf{h}) - w(\mathbf{x})\}^2 / p_k(\mathbf{h})$$

③次式により, 各ヴァリオグラムを重ね合わせて (重み付き平均をとって) 全体のヴァリオグラム $\gamma(\mathbf{h})$ とする.

$$\gamma(\mathbf{h}) = \sum_{k=1}^m p_k(\mathbf{h}) \gamma_k(\mathbf{h}) / \sum_{k=1}^m p_k(\mathbf{h})$$

この手順の結果得られた上式の $p_k(\mathbf{h})$ の $\mathbf{h}$ に関する和 $p_k^\Sigma = \sum_{\mathbf{h}} p_k(\mathbf{h})$ は, 部分計算でグループ k のヴァリオグラムを計算するときの対の数の総和を表し, 次式で与えられる.

$$p_k^\Sigma = (q_m + 1)q_m / 2 \quad (k \leq r_m)$$

$$p_k^\Sigma = q_m(q_m - 1) / 2 \quad (k > r_m)$$

データを m 個のグループに分けてヴァリオグラムを計算したときの対の数を P_m とすると, 次式が得られる.

$$P_m = \sum_{k=1}^m p_k^\Sigma$$

$$= \sum_{k=1}^{r_m} (q_m + 1)q_m / 2 + \sum_{k=r_m+1}^m q_m(q_m - 1) / 2$$

$$= \{(q_m + 1)r_m + (q_m - 1)(m - r_m)\}q_m / 2$$

$$= q_m \{mq_m + r_m - (m - r_m)\} / 2$$

$$= q_m \{mq_m - m + 2r_m\} / 2 \quad (5)$$

したがって, グループ化部分計算で用いられる対の数 P_m は次式のように近似できる [注2].

$$P_m \approx P/m$$

第2図は, 各部分計算において採用された対が, ランダムに並び替える前のデータの並びでどこに位置していたかを模式的に示している. 各図において, 左下の三角形内で打点されたセルが計算に採用された対を, 右上の三角形内で打点されたセルがランダムに並べ替えられる前の並びにおけるそれらの位置を示している. 一部データ部分計算と先頭優先部分計算では, 対がランダムに採用されていないことが一目瞭然である. これに対して, 近隣順序部分計算とグループ化部分計算では対がランダムに採用されているように見える. しかし, 近隣順序部分計算では各データとも4回採用されており, クループ化部分計算では各データとも3ないし4回採用されていて, 対が全体からランダムに採用されているわけではない.

3. 計算の進行と収束

今回使用したデータのヴァリオグラムの特性を理解するために, 完全計算で得られた全方向実験ヴァリオグラムを第3図に, 直交実験ヴァリオグラムを方向実験ヴァリオグラムから得られたレンジとともに第4図に示す. なお, 本論文で図示されているヴァリオグラムはすべて実験ヴァリオグラムなので, 以下では単に「ヴァリオグラム」と記す. 第3図には, 完全計算に較べて対の数が約1/100である部分計算により得られた全方向ヴァリオグラムも合わせて示してある. また, 第5図には, やはり対の数が完全計算に較べて約1/100の条件で行った近隣順序部分計算で得ら

れた直交ヴァリオグラムが示してある。

まず初めに、部分計算の結果が、計算の進行に伴って、どのように完全計算の結果に近づくかを見る。部分計算の進行は、完全計算における対の数に対する部分計算の途中段階における対の数の比で定義する。すなわち、式(2)~(4)の P_n と完全計算の対の数 $P=N(N-1)/2$ を使って、 P_n/P で表す。そこで、式(2)~(4)の n が各部分計算における「現段階」を示しているとし、それを $(n-1)$ で置いた場合を「1段階前」と表現する。なお、グループ化部分計算は、計算を進行させながらグループ数を逐次減少させるという操作ができないので、ここでは検討の対象としない。

部分計算で得られたヴァリオグラムと完全計算で得られたヴァリオグラムの類似性は、両者の間の相関係数で評価する。いま2つのヴァリオグラム1と2があって、 h で結びつけられた両者のヴァリオグラム値、すなわち $\gamma_1(h)$ と $\gamma_2(h)$ を対として相関係数を求めると、両者が完全に一致するならば、その値は1となる。したがって、2つのヴァリオグラムの間の相関係数が高いほど、両者は似ていると判断できる。なお、相関係数の計算に当たっては、一方のヴァリオグラムは h で値が与えられているのに他方のヴァリオグラムでは与えられていない場合は、相関係数の計算からはずす。ただし、対の数 $p_1(h)$ と $p_2(h)$ の相関係数を求めるときは、ヴァリオグラム値が与えられていない方の対の数は0として、相関係数の計算に使う。この相関係数を、以下の理由により「単純相関係数」と呼ぶ。ヴァリオグラムの各値は、それぞれ異なる対の数から求められている。これに対して、単純相関係数はヴァリオグラム値が求めら

第3図. 全方向実験ヴァリオグラムとそれらに当てはめられた球モデル (実線). 白抜き丸が完全計算で得られた実験ヴァリオグラムで、黒塗りの三角、黒塗りの菱形、十字、X字が、それぞれ一部データ部分計算、先頭優先部分計算、近隣順部分計算、グループ化部分計算で得られた実験ヴァリオグラムである。各部分計算とも対の数が完全計算と較べて約1/100となる条件で10回行い、完全計算の実験ヴァリオグラムとの相関が最も低かった場合が示してある。3本見える曲線のうち、上は完全計算、近隣順部分計算、グループ化部分計算の球モデルが重なっている。中と下の曲線は、それぞれ先頭優先計算と一部データ計算の球モデルである。

第4図. 完全計算で得られた直交実験ヴァリオグラムと方向実験ヴァリオグラムで求められたレンジ (白丸付き折線). 両軸とも表示範囲は±5kmで、濃淡で表現されている実験ヴァリオグラム値の境界は、濃から淡へ5000, 2000, 1000m²である。

第5図. 近隣順部分計算で得られた直交実験ヴァリオグラムと方向実験ヴァリオグラムから求められたレンジ (白丸付き折線). 計算は対の数が完全計算と較べて約1/100となる条件で10回行い、完全計算の実験ヴァリオグラムとの相関が最も低かった場合が示してある。両軸の表示範囲と濃淡で表現されている実験ヴァリオグラム値の境界は第4図と同じ。

第6図. 部分計算の進行に伴う中途段階と完全計算の実験ヴァリオグラムの間の相関係数 (r) の変化 (全方向実験ヴァリオグラムの場合). 黒塗りの丸, 三角, 菱形の点列は, それぞれ一部データ部分計算, 先頭優先部分計算, 近隣順部分計算を示し, 白抜き点列は, 黒塗りの点列のうち進行度が0と0.1の間を横軸方向に10倍に伸ばしてある (点はすべてを記すと重なるので, 間引いてある). 横軸の計算の進行度は, 完全計算の対の数に対する中途段階の対の数の割合. 縦軸は, $(1+r)/(1-r)$ の値を対数目盛りで示している (目盛りの印位置は100を底とする). 右側には, その値に対応する相関係数が示してあるが, 小数点以下に並ぶ9の数は添字の数字で示してある. なお, 0と0.9₁ (=0.9) の間の目盛り位置は0.5.

れたときの対の数が考慮されていないので, 得られた値は結果として対の数が少ないヴァリオグラム値が相対的に強調されている (第3図では, ずれが3km以上の部分). それを防ぐためには, 各ヴァリオグラム値でそれぞれの対の数を重みとして考慮に入れる必要がある. つまり, $\gamma(h_1)$ と $\gamma(h_2)$ を求めるときの対の数はそれぞれ $p(h_1)$ と $p(h_2)$ であり, 通常 $p(h_1) \neq p(h_2)$ であるから, 相関係数を求めるときこの対の数を考慮に入れるべきである. ただし, 相関係数を求めようとする2つ計算結果の $p_1(h)$ と $p_2(h)$ は同じ値にはない. そこで, $p_1(h)$ と $p_2(h)$ の和を距離ベクトル h における対の数とした [注3]. この相関係数を, 以下, 「重み付き相関係数」と呼ぶ.

部分計算の進行に伴う部分計算と完全計算のヴァリオグラムの間の重付き相関係数の変化を, 全方向ヴァリオグラムは第6図に, 直交ヴァリオグラムは第7図に示す. いずれのヴァリオグラムも, 相関係数は, 部分計算の開始後急速に増加して, 進行度が約1%で0.9に達し, その後は緩やかな増加傾向を示す (進行度が1に近づくとき急激な増加傾向を示すが, この縦軸の目盛りでは $r=1$ で無限大に発散する). ただし, 全方向ヴァリオグラム (第6図) と直交ヴァリオグラム (第7図) を較べた場合, 直交ヴァリオグラムの相関係数が滑らかな増加傾向であるのに対し, 全方向ヴァリオグラムの相関係数は細かい増減を繰り返しながらの増加傾向である. 部分計算の方法による違いを見ると, いずれのヴァリオグラムにおいても, 近隣順部分計算

第7図. 部分計算の進行に伴う中途段階と完全計算の実験ヴァリオグラムの間の相関係数 (r) の変化 (直交実験ヴァリオグラムの場合). 凡例や軸は第6図と同じ. ただし, 黒塗りの丸と三角はほとんど重なっている.

の相関係数があらゆる進行度で高い値を示しているのに対し, 一部データ部分計算と先頭優先部分計算の間では差がほとんど認められない. 近隣順部分計算の優位性は, 直交ヴァリオグラムで特に顕著で, 他に対して小数点直下に並ぶ9の数が1個は多い.

4. 計算結果とその評価

第6図と第7図は, 各部分計算とも1回しか実行していないために相関係数が各進行度でどの程度変動するかが分らない. そこで, 対の数が完全計算の場合 ($39520495 \approx 40 \cdot 10^6$ 対) の約1/10, 1/100, 1/1000 (それぞれ進行度で10%, 1%, 0.1%) の場合について, データの並びを乱数で10回変えて, 極ヴァリオグラム (全方向と10°毎の方向ヴァリオグラム) と直交ヴァリオグラムを部分計算で求めるとともに, それぞれのヴァリオグラムについて, 完全計算のヴァリオグラムとの間の相関係数を求めた. 第1表には, 得られた10個の相関係数のうちの最小値が示してある. ところで, 部分計算を適用しなければならないほどデータが多量に存在する場合は, 基本的に完全計算が実行できない. したがって, 完全計算の結果と比較して部分計算の精度を評価するのではなく, 他の方法で評価する必要がある. そのような場合を想定して, 各部分計算における10回の結果の間における45の相関係数のうちの最小値も第1表には示してある. また, ヴァリオグラムの重要なパラメーターであるレンジについても, ヴァリオグラムと同様に相関係数を求め, その値が示してある. なお, 第1表にはグループ化部分計算の結果も示してある.

部分計算に要した時間は, 進行度が10%の場合極ヴァリオグラムが446~742秒で直交ヴァリオグラムが5552~6611秒, 1%の場合それぞれ62~80秒と531~588秒, 0.1%の場合それぞれ15~36秒と51~150秒であった. これらの値を完全計算の場合 (それぞれ5422秒と55859秒)

第1表. 完全計算 (サンプル数=8891, 対の数=39520495 $\approx 40 \cdot 10^6$) と部分計算で得られたヴァリオグラムとの類似性を示す相関係数と計算時間.

計算法	レンジ	全方向ヴァリオグラム		直交ヴァリオグラム	
		単純	重み付き	単純	重み付き
一部データ	0.9992	0.9944	0.9998	0.9886	0.9975
試料数=2963	0.9973	0.9867	0.9970	0.9701	0.9929
対の数=4388203	計算時間	742 秒 (12:22)		6611 秒 (1:50:11)	
先頭優先	0.9941	0.9913	0.9968	0.9898	0.9935
先頭数=444	0.9937	0.9762	0.9943	0.9593	0.9874
対の数=390214	計算時間	646 秒 (10:46)		5741 秒 (1:35:41)	
近隣順	1.0000	0.9991	1.0000	0.9993	0.9999
近隣数=444	0.9999	0.9984	1.0000	0.9983	0.9998
対の数=391160	計算時間	613 秒 (10:13)		6465 秒 (1:47:45)	
グループ化	1.0000	0.9989	1.0000	0.9992	0.9999
グループ数=10	1.0000	0.9976	1.0000	0.9983	0.9998
対の数=390808	計算時間	446 秒 (7:46)		5552 秒 (1:32:32)	
一部データ	0.9908	0.9470	0.9917	0.9577	0.9853
試料数=889	0.9754	0.9066	0.9857	0.9087	0.9660
対の数=394716	計算時間	62 秒 (1:02)		538 秒 (8:58)	
先頭優先	0.9629	0.9272	0.9744	0.8639	0.9200
先頭数=44	0.9225	0.8617	0.9287	0.6551	0.8337
対の数=390214	計算時間	80 秒 (1:20)		572 秒 (9:32)	
近隣順	0.9998	0.9885	0.9999	0.9892	0.9990
近隣数=44	0.9995	0.9661	0.9997	0.9768	0.9980
対の数=391160	計算時間	71 秒 (1:11)		588 秒 (9:48)	
グループ化	0.9998	0.9882	0.9999	0.9913	0.9990
グループ数=100	0.9997	0.9654	0.9996	0.9832	0.9979
対の数=390808	計算時間	74 秒 (1:14)		531 秒 (8:51)	
一部データ	0.9256	0.9537	0.9713	0.8808	0.9459
試料数=270	0.8789	0.8905	0.9349	0.6964	0.8821
対の数=36046	計算時間	15 秒		51 秒	
先頭優先	-0.0159	0.3350	0.5529	0.0721	0.3182
先頭数=4	-0.2002	-0.2048	0.2264	-0.2319	0.0431
対の数=35554	計算時間	16 秒		57 秒	
近隣順	0.9980	0.9629	0.9986	0.9434	0.9883
近隣数=4	0.9945	0.9132	0.9968	0.9050	0.9751
対の数=35560	計算時間	15 秒		58 秒	
グループ化	0.9974	0.9404	0.9979	0.9554	0.9887
グループ数=1000	0.9943	0.8921	0.9949	0.9089	0.9773
対の数=35128	計算時間	36 秒		150 秒 (2:30)	

各計算法・ヴァリオグラムとも, 上段が完全計算と部分計算の結果の間の相関係数, 下段が部分計算による結果間の相関係数. なお, 部分計算は異なる乱数で 10 回行い, 相関係数は得られたすべての組合せのうちの最低値を示す.

ヴァリオグラムの単純相関係数は座標で結びつけられたヴァリオグラム値を対とし求め, 重み付き相関係数はこれにヴァリオグラムを計算するときの対の数を重みとして加味して求めた.

全方向ヴァリオグラムの計算時間欄の数値は, 18 方向のヴァリオグラムの計算も含む極座標ヴァリオグラムの計算時間である. なお, 完全計算の計算時間は, 極座標ヴァリオグラムが 5422 秒 (1:30:02), 直交ヴァリオグラムが 55859 秒 (15:30:59)であった.

と較べると、進行度が10%と1%の場合はそれぞれ1/10と1/100より多少長いもののほこの値となっているのに対し、0.1%の場合は1/1000よりかなり長い。特に、グループ化部分計算でこの傾向が強い。実際の計算時間が、このように期待される値より長くなるのは、対の数に関係しない計算の部分があるためである。

第1表に示されている相関係数は、10回行なった部分計算から得られたヴァリオグラムと完全計算のヴァリオグラムの間の相関係数のうちの最小の値である。すなわち、ここに示された値より悪い結果が出る確率は、大雑把には1/10以下と推定される。近隣順部分計算とグループ化部分計算では、完全計算に較べて1/10の対を使う場合重付き相関係数で0.999以上、1/100の対を使う場合でも0.99以上の類似性が保証される。これに対して、1/1000の対を使う計算で保証される類似性は、相関係数で0.9である。したがって、1/1000の対を使う計算は、計算時間が期待ほど短縮されないのに精度が低下するので、推奨できない。そこで、対の数が完全計算の1/1000の場合は評価の対象としない。

次に、10回行なった部分計算の結果の間の相関係数を見る。完全計算の結果との間の相関係数と同様に、近隣順部分計算の結果の相関係数が高く、全方向ヴァリオグラムでも直交ヴァリオグラムでも重み付き相関係数では0.9以上の値が出ている。また、レンジの値の相関係数は0.99を越えている。グループ化部分計算の結果も、近隣順部分計算の結果と較べて多少値が低いものの、十分に高い相関係数が出ている。これに対して、一部データ部分計算と先頭優先部分計算の結果は、一部で相関係数が0.9を切っており、

十分な精度が出ているとは言い難い。このことは、完全計算の結果と相関の高い結果を出す部分計算は、その計算方法で出された複数の結果の間でも高い相関性を示すということであり、異なるデータ配列で出た結果の間の相関が高ければ、その手法で出された部分計算の結果は完全計算と高い相関を示すと推定できることを示唆している。

ヴァリオグラムの値は、許容範囲を変えると変わる[注1]。例えば、今回の例の場合、許容範囲を0.1, 0.2, 0.3kmと変えて完全計算を実施した場合のヴァリオグラムのレンジの間の相関係数は、許容範囲が0.1と0.2kmの間が0.99997, 0.1と0.3kmの間が0.99986, 0.2と0.3kmの間が0.99996である。また、全方向ヴァリオグラム間の重付き相関係数は、それぞれ0.99995, 0.9998, 0.99992, 直交ヴァリオグラム間の重付き相関係数は、それぞれ0.99986, 0.9995, 0.9998である。これらの値と、上述の部分計算で得られた値を比較すると、近隣順部分計算は、1/10の対を使う場合完全計算と同程度以上の精度が出ており、1/100の対を使う場合でもほぼ同程度の精度が出ていると結論できる。グループ化部分計算も0.9998の相関係数を示す(第1表)ので、近隣順部分計算と比較して遜色がない。

地球統計学では扱うデータは定常性(stationarity)を満たしている必要がある。しかし、実際にそれを検証することは極めて難しく、通常は定常性が満たされていると仮定して作業を進めている。ところが、上に紹介したような複数の部分計算でヴァリオグラムを求めると、ある程度定常性の検証が可能となる。すなわち、得られた複数のヴァリオグラムを比較して、それらの類似性が高ければ、定常性が満足されている可能性が高い。厳密にいうと、類似性が低ければ「定常性が満足されている」という前提が棄却されるのに対し、類似性が高いとこの前提が棄却されないの、「定常性が満足されている」と見なすことができるということである。これは、平均値の差の検定で、「2つの母

第8図. 近隣順部分計算の進行による実験ヴァリオグラムの間およびそれらの計算に用いられた対の数の間の相関係数の変化(全方向実験ヴァリオグラムの場合). 黒丸の点群が中途段階とその1段階前に得られた実験ヴァリオグラムの間、白抜き記号の点群は上が中途段階とその1段階前の対の数の間、下が中途段階と完全計算の実験ヴァリオグラムの間、中が中途段階と完全計算の対の数の間の相関係数を示す。なお、黒丸の点群の上限は、それを隠している白丸の点群で下から2列目の点列の辺りを通る。横軸と縦軸は第6図と同じ。

第9図. 近隣順部分計算の進行による実験ヴァリオグラムの間およびそれらの計算に用いられた対の数の間の相関係数の変化(直交実験ヴァリオグラムの場合). 凡例や軸は第8図と同じ。

集団の平均値は等しい」という前提が棄却されないので、「2つの母集団の平均値は等しい」と見なすということと同じ論理である。なお、定常性の検証には、部分計算の手法としては劣っていた一部データ部分計算や先頭優先部分計算の方が、データや対を偏って使っているのに向いているのかもしれない。これに関する議論は今後の課題であろう。

5. 計算の打ち切り

前章では、多量のデータがあってヴァリオグラムの完全計算が不可能な場合を想定して、最初に計算に使う対の数を設定する部分計算を説明した。しかし、実際の計算では、十分な精度が出たところで、計算を打ち切る方が合理的である。そこで、部分計算で最も精度が高い結果を示し、計算の途中打ち切りが可能な近隣順部分計算について、計算の途中打ち切りを行なう判断基準について検討する。この判断のために、前述した相関係数を、中途段階と完全計算のヴァリオグラムの間 (r_{VPC} と略記: V = Variogram, PC = between Partial and Complete), 中途段階とその1段階前のヴァリオグラムの間 (r_{VPP} : PP = between Partial and Partial), 中途段階と完全計算の対の数の間 (r_{PPC} : P = Pairs), 中途段階とその1段階前の対の数の間 (r_{PPP}) で求める。これらの相関係数の変量を一々記述すると長くなるので、以下では括弧内に示した略号で記す。

近隣順部分計算における r_{VPC} と r_{VPP} の変化を、計算の進行の関数として、全方向ヴァリオグラムは第8図に、直交ヴァリオグラムは第9図に示す。また、両図には r_{PPC} と r_{PPP}

も示してある。第8図に示されている全方向ヴァリオグラムでは、各相関係数とも計算の進行とともに増加する傾向を示す。しかし、 r_{PPC} を除いて、単調に増加するのではなく途中で増減を繰り返している。特に、 r_{VPP} と r_{PPP} は増減の変動が激しく、図では点列ではなく点群をなしている。一方、第9図に示されている直交ヴァリオグラムの場合も傾向は全方向ヴァリオグラムと同様であるが、その増減の変動はかなり小さい。

ヴァリオグラムの計算を打ち切る段階を判断するために、 r_{VPC} を r_{VPP} あるいは r_{PPP} と計算の進行度で結びつけて、散布図を描いた(第10図と第11図)。第10図に示されている全方向ヴァリオグラムの場合、 r_{VPC} は r_{VPP} あるいは r_{PPP} が増加しても、0.9からはほとんど増加していない。一方、第11図に示されている直交ヴァリオグラムの場合、 r_{VPC} は r_{VPP} あるいは r_{PPP} の増加にともなって徐々に増加しており、小数点以下に並ぶ9の数でみると、その数が半分となる値は保証されているといえる。すなわち、 r_{VPC} は r_{VPP} あるいは r_{PPP} が0.99ならば0.9以上、0.9999ならば0.99以上、一般化して表記すると、 r_{VPP} あるいは r_{PPP} が 0.9_s (s は小数点以下に並ぶ9の数) ならば r_{VPC} は $0.9_{s/2}$ (小数点以下に並ぶ9の数が s の1/2) 以上である。

第1表によると、極ヴァリオグラムの計算時間は直交ヴァリオグラムに較べて約1/10である。したがって、極ヴァリオグラムは完全計算を行なって、その結果にもとづいて、直交ヴァリオグラムの打ち切り時を判断することが考えられる。また、第6図や第10図で明らかなように、 r_{VPC} は計算が進行しても変動が激しく、徐々に向上する傾向は示さない。このことは第10図に示されている r_{VPC} の点列が、乱数の系列毎に変わる可能性を示唆している。そこで、10の乱数系列で極ヴァリオグラムを計算して、それぞれにおける

第10図. 近隣順部分計算における中途段階の実験ヴァリオグラムおよびその計算に用いられた対の数と完全計算の実験ヴァリオグラムの相関係数 (r) の関係 (全方向実験ヴァリオグラムの場合)。横軸は、中途段階とその一段前の実験ヴァリオグラムの間の相関係数 (黒丸) およびそれらの実験ヴァリオグラムの計算に用いられた対の間の相関係数 (白丸) を表し、縦軸は中途段階の部分計算と完全計算で得られた全方向実験ヴァリオグラムの間の相関係数を表す。両軸は $(1+r)/(1-r)$ の値を対数目盛りで示し、図の右と上の目盛りはその値に対応する相関係数を示す (詳細は第6図と同じ)。

第11図. 近隣順部分計算における中途段階の実験ヴァリオグラムおよびその計算に用いられた対の数と完全計算の実験ヴァリオグラムの相関係数 (r) の関係 (直交実験ヴァリオグラムの場合)。凡例や軸は第10図と同じ。

r_{VPC} を求め、それらと直交ヴァリオグラムの r_{VPC} を計算の進行度で結びつけて、散布図を描いた(第12図)。第12図では、系列の異なる乱数から求められた全方向ヴァリオグラムの r_{VPC} が変動しながら重なり合っていて、それぞれの r_{VPC} を個々の乱数系列毎に把握することはできない。つまり、乱数の系列が変わると、この図で示されている程度の変動はあるということが分る。図では、全方向ヴァリオグラムの r_{VPC} の値に対して直交ヴァリオグラムのその値が、小数点以下の9の数にして2個少なくなるときを破線で示してある。図に示されている散布図の各点はこの線より上にあるので、直交ヴァリオグラムの r_{VPC} の値は、全方向ヴァリオグラムの r_{VPC} の値と較べて、小数点以下の9の数にして2個少ない値よりは高いと結論できる。

上述の結果から、次の結論が得られる。例えば、いま完全計算の結果との相関係数(r_{VPC})が0.99以上になる部分計算の直交ヴァリオグラムを望むならば、中途段階とその1段階前のヴァリオグラムの間の相関係数(r_{VPP})が小数点以下に並ぶ9の数で2倍、すなわち0.9999以上になったところで計算を打ち切れればよい。また、極ヴァリオグラムの r_{VPC} で直交ヴァリオグラムの部分計算に必要な対の数は次の方法で推定することができる。例えば、直交ヴァリオグラムの r_{VPC} で0.9以上を望むならば、極ヴァリオグラムの r_{VPC} は小数点のあとに並ぶ9の数が2個多い0.999が必要なので、その進行度に相当する対の数を直交ヴァリオグラムの計算に適用すればよい。

6. 結論

ヴァリオグラムを得るための計算時間は、データ数の2乗に比例して長くなる。この欠点を避けるために、1) データの一部を使う一部データ部分計算、2) データは全部使うが対の相手は先頭部分しか使わない先頭優先部分計算、3) 各データとも近隣のデータとの対しか使わない近隣順部分計算、4) ほぼ同数のデータからなるグループに分けて、各グループで得られたヴァリオグラムを重ね合わせて全体のヴァリオグラムとするグループ化部分計算を実行し、結果を完全計算の結果と比較して、精度を評価した。その結果、一部データ部分計算と先頭優先部分計算の精度は低く、近隣順部分計算とグループ化部分計算の精度は高かった。約9000個のデータからなる今回の例では、対の数が完全計算の約1/100の場合、完全計算で得られたヴァリオグラムとの相関係数は0.99以上であった。計算を途中で打ち切る場合、現段階と1段階前のヴァリオグラムの相関係数が0.9₉(sは小数点以下の9の数)ならば、現段階と完全計算の直交ヴァリオグラムの間の相関係数は0.9_{9/2}(小数点以下の9の数が約半分)が保障される。

上述した部分計算は、日本情報地質学会のホームページからダウンロードできるVBAVariogram (MS-Excel/VBAで計算するヴァリオグラム：正路, 2002) に既に組み込んである。

第12図. 近隣順部分計算で得られた実験ヴァリオグラムで進行度を媒介変数として結びつけたときの全方向実験ヴァリオグラムの中途段階と完全計算の実験ヴァリオグラムの間の相関係数(横軸)と直交実験ヴァリオグラムの中途段階と完全計算の実験ヴァリオグラムの間の相関係数(縦軸)との関係。全方向実験ヴァリオグラムの計算は異なる乱数で10回行った。両軸の目盛りは第10図と同じ。一点鎖線と破線は、それぞれ縦軸の値と横軸の値の比が真数で1/1と1/100の場合を示す(後者の場合、全方向実験ヴァリオグラムの相関係数の小数点以下の9の数が直交実験ヴァリオグラムのそれより約2個多い)。

謝辞 東京大学海洋研究所の田村千織氏と沖野郷子氏は解析用のデータを提供して下さい。中国職業能力開発大学校講師の小林文美子修士は草稿を読んで貴重な意見を下さった。記して、感謝の意を表する。

文 献

- 正路徹也(2002) MS-Excel/VBAによるヴァリオグラムの計算. 情報地質, vol.13, no.1, pp.9-21.
 正路徹也・小池克明(2007) 講座「地球統計学」ヴァリオグラム—データの空間的連続性の解析. 地熱学会誌, vol.29, no.3, pp.125-140.

注1: 式(1)で分子の $\sum \{w(\mathbf{x} + \mathbf{h}) - w(\mathbf{x})\}^2$ は、2点間の距離が $\mathbf{h}$ である対のみを取り出して、両点における測定値の差の二乗和を求めることを意味する。ただし、実際の実験ヴァリオグラムの計算に当たっては、距離 $\mathbf{h}$ を離散的に与えるとともに許容範囲 $\Delta \mathbf{h}$ を定義して、 $\mathbf{h} \pm \Delta \mathbf{h}$ の距離をもつ対は距離が $\mathbf{h}$ であると見なして $\gamma(\mathbf{h})$ の算出に加える(例えば、正路・小池, 2007)。測点が1次元の直線(曲線ならば、変換された直線)上に並んでいる場合、 $\mathbf{h}$ を定数 d の整数倍で与えるとともに、 $\Delta \mathbf{h}$ を $d/2$ とすると、すべての対を必ず1回そしてただ1回ヴァリオグラムの計算に使うことができる。この場合、当然のことながら $\mathbf{h} \pm \Delta \mathbf{h}$ は同じ形状で隙間なく排他的に配置されている。このように $\mathbf{h} \pm \Delta \mathbf{h}$ を隙間なく排他的に配置でき

るときは、ヴァリオグラムの計算に使う対の数の総和は $p^{\Sigma} = \sum p(\mathbf{h})$ は $N(N-1)/2$ である。これに対して、データが 2^n 次元以上の空間に散らばっている場合は、 $\mathbf{h} \pm \Delta \mathbf{h}$ を同じ形状にして隙間なく排他的に配置することはできない。例えば、2次元の場合、内孤半径が $h - \Delta h$ 、外孤半径が $h + \Delta h$ 、開角が $\pm \Delta \theta$ の扇形は隙間なく排他的に配置できるが、形状が $\mathbf{h}$ で変わる。また、中心が $\mathbf{h}$ で半径が Δh の円は、形状は同じだが、隙間なく排他的に配置することができない。このため、例えば $\mathbf{h}_1 \pm \Delta \mathbf{h}$ と $\mathbf{h}_2 \pm \Delta \mathbf{h}$ で与えられる範囲が重なり、ある対が $\gamma(\mathbf{h}_1)$ と $\gamma(\mathbf{h}_2)$ の両方の計算に使われる。その結果として、同じ対が $p(\mathbf{h}_1)$ と $p(\mathbf{h}_2)$ の両方に数えられて、 p^{Σ} は $N(N-1)/2$ より大きくなる。ただし、この現象を、今回の部分計算で対の数が完全計算の $1/100$ となる場合における完全計算と部分計算における p^{Σ} の値の比、あるいは部分計算同士の p^{Σ} の値の比で見ると、前者で $1/100$ から最も離れた値は一部データ部分計算が $1/99.96$ 、先頭優先部分計算が $1/101.3$ 、近隣順部分計算が $1/101.0$ 、グループ化部分計算が $1/101.2$ であり、後者で $1/1$ から最も離れた値はそれぞれ 1.0009 , 1.002 , 1.0003 , 0.999 である（前者で比が $1/100$ に比して一部部分計算では常に大きく他の部分計算では常に小さくなるが、それは近似式の導き方とそれから n を決めるときの丸め誤差による）。したがって、この現象は、ここでの議論の本質にほとんど影響を与えない。

注2：グループ化部分計算において、 N を m で割った余り r_m は N の増加とともに、 0 から $(m-1)$ までの変化を繰り返す。そこで、 r_m が繰返しどの範囲にあるかを明確にするために、 q_m が一定の範囲を範囲 q_m と定義する。この定義に従うと、式(5)の P_m は q_m と r_m の関数となるので、以下 $P_m(q_m, r_m)$ と書く。範囲 q_m において、 N が 1 増加すると r_m も 1 増加する。これを数式で表現すると、 $\Delta N = 1$ ならば $\Delta r_m = 1$ 、すなわち $\Delta r_m / \Delta N = 1$ である。したがって、範囲 q_m においては、式(5)から、

$$\begin{aligned} \frac{\Delta P_m(q_m, r_m)}{\Delta N} &= \frac{P_m(q_m, r_m+1) - P_m(q_m, r_m)}{\Delta r_m} \cdot \frac{\Delta r_m}{\Delta N} \\ &= \frac{q_m\{mq_m - m + 2(r_m+1)\}}{2} \cdot \frac{q_m\{mq_m - m + 2r_m\}}{2} \\ &= q_m \end{aligned}$$

が得られる。さらに N が 1 増えて隣の範囲に入る、すなわち範囲 q_m から範囲 (q_m+1) へ移る場合を考えると、商は q_m から (q_m+1) へ、余りは $(m-1)$ から 0 へ変わる。このとき、

$$\begin{aligned} \Delta P_m(q_m, r_m) &= P_m(q_m+1, 0) - P_m(q_m, m-1) \\ &= \frac{(q_m+1)\{m(q_m+1) - m\}}{2} \\ &\quad - \frac{q_m\{mq_m - m + 2(m-1)\}}{2} \\ &= \frac{mq_m^2 + 2mq_m + m - mq_m - m}{2} \end{aligned}$$

$$\begin{aligned} &= \frac{mq_m^2 - mq_m + 2mq_m - 2q_m}{2} \\ &= q_m \end{aligned}$$

である。したがって、ここでも、

$$\frac{\Delta P_m}{\Delta N} = q_m$$

である。すなわち、商が q_m から (q_m+1) になるまでの範囲において $\Delta P_m(q_m, r_m) / \Delta N = q_m$ である。また、 N の増加に伴って q_m は階段状に増加して減少することはないので、 N の増加に伴って P_m は単調に増加する。

$N = mq_m + r_m$ であることを考慮すると、全データを使う場合の対の数 P (これをグループの数が 1 と考えると P_1) は次式で与えられる。

$$\begin{aligned} P_1(q_m, r_m) &= P \\ &= \frac{(mq_m + r_m)(mq_m + r_m - 1)}{2} \\ &= \frac{m^2q_m^2 + 2mq_m r_m - mq_m + r_m^2 - r_m}{2} \end{aligned} \quad (\text{A1})$$

したがって、 m 個のグループに分けた場合の対の数と全データを使った場合の対の数との比、すなわち $t_m(q_m, r_m) = P_m(q_m, r_m) / P_1(q_m, r_m)$ は、式(5)と(A1)より次式で与えられる。

$$\begin{aligned} t_m(q_m, r_m) &= \frac{P_m(q_m, r_m)}{P_1(q_m, r_m)} \\ &= \frac{q_m(mq_m - m + 2r_m)}{m^2q_m^2 + 2mq_m r_m - mq_m + r_m^2 - r_m} \end{aligned} \quad (\text{A2})$$

ところで、範囲 q_m においてデータ総数 N が 1 増えた場合の $t_m(q_m, r_m)$ の増加量 $\Delta t_m(q_m, r_m)$ は次の式 (A3) で与えられる。ただし、ここでは $r_m \geq 0$ で $r_m \leq m-2$ とし、 $r_m = m-1$ の場合は N を 1 増加すると範囲が隣へ移るので別に考える。

$$\begin{aligned} \Delta t_m(q_m, r_m) &= \Delta \{P_m(q_m, r_m) / P_1(q_m, r_m)\} \\ &= \frac{P_m(q_m, r_m+1)}{P_1(q_m, r_m+1)} - \frac{P_m(q_m, r_m)}{P_1(q_m, r_m)} \\ &= \frac{P_m(q_m, r_m+1)P_1(q_m, r_m) - P_m(q_m, r_m)P_1(q_m, r_m+1)}{P_1(q_m, r_m+1)P_1(q_m, r_m)} \\ &= \frac{4A}{4P_1(q_m, r_m+1)P_1(q_m, r_m)} \end{aligned} \quad (\text{A3})$$

ここで、式(A3)の分子の $4A$ は次の式(A4)で与えられる。

$$\begin{aligned} 4A &= 2P_m(q_m, r_m+1) \cdot 2P_1(q_m, r_m) - 2P_m(q_m, r_m) \cdot 2P_1(q_m, r_m+1) \\ &= q_m\{mq_m - m + 2(r_m+1)\} \{m^2q_m^2 + 2mq_m r_m - mq_m + r_m^2 - r_m\} \\ &\quad - q_m\{mq_m - m + 2r_m\} \\ &\quad \times \{m^2q_m^2 + 2mq_m(r_m+1) - mq_m + (r_m+1)^2 - (r_m+1)\} \\ &= q_m\{mq_m - m + 2r_m + 2\} \{m^2q_m^2 + 2mq_m r_m - mq_m + r_m^2 - r_m\} \\ &\quad - q_m\{mq_m - m + 2r_m\} \\ &\quad \times \{(m^2q_m^2 + 2mq_m r_m - mq_m + r_m^2 - r_m) + (2mq_m + 2r_m)\} \\ &= q_m(mq_m - m + 2r_m)(m^2q_m^2 + 2mq_m r_m - mq_m + r_m^2 - r_m) \\ &\quad + 2q_m(m^2q_m^2 + 2mq_m r_m - mq_m + r_m^2 - r_m) \\ &\quad - q_m(mq_m - m + 2r_m)(m^2q_m^2 + 2mq_m r_m - mq_m + r_m^2 - r_m) \\ &\quad - 2q_m(mq_m - m + 2r_m)(mq_m + r_m) \end{aligned}$$

$$\begin{aligned}
 &= 2q_m(mq_m + r_m)(mq_m + r_m - 1) \\
 &\quad - 2q_m(mq_m - m + 2r_m)(mq_m + r_m) \\
 &= 2q_m(mq_m + r_m)\{(mq_m + r_m - 1) - (mq_m - m + 2r_m)\} \\
 &= 2q_m(mq_m + r_m)(m - r_m - 1) \quad (A4)
 \end{aligned}$$

定義により, $q_m \geq 1$, $m \geq 2$ ($m = 1$ だと, グループ分けを行わないので, ここでは無意味), $r_m \geq 0$, $m - r_m \geq 2$ であるから, 式(A4)のAは常に正である. また, 式(A3)の分母はデータ総数が N と $(N+1)$ の場合の対の数の積であるから, これも常に正である. したがって, 範囲 q_m において $\Delta t_m(q_m, r_m)$ は常に正である. 次に, 範囲 q_m から範囲 $(q_m + 1)$ へ移る場合, すなわち商が q_m から $(q_m + 1)$ へ, 余りが $(m - 1)$ から0へ変わる場合を考える. そのための準備として, 式(A2)で $r_m = 0$ とすると, 次の式(A5)が得られる.

$$\begin{aligned}
 t_m(q_m, 0) &= \frac{q_m(mq_m - m)}{m^2 q_m^2 - mq_m} \\
 &= \frac{mq_m(q_m - 1)}{mq_m(mq_m - 1)} \\
 &= \frac{q_m - 1}{mq_m - 1} \quad (A5)
 \end{aligned}$$

式(A5)で q_m を $(q_m + 1)$ と置くと, 次の式(A5')が得られる.

$$\begin{aligned}
 t_m(q_m + 1, 0) &= \frac{(q_m + 1) - 1}{m(q_m + 1) - 1} \\
 &= \frac{q_m}{mq_m + m - 1} \quad (A5')
 \end{aligned}$$

一方, 式(A2)で $r_m = m - 1$ とすると, 次の式(A6)が得られる.

$$\begin{aligned}
 t_m(q_m, m - 1) &= \frac{q_m\{mq_m - m + 2(m - 1)\}}{m^2 q_m^2 + 2mq_m(m - 1) - mq_m + (m - 1)^2 - (m - 1)} \\
 &= \frac{q_m(mq_m + m - 2)}{m^2 q_m^2 + 2m^2 q_m - 2mq_m - mq_m + m^2 - 2m + 1 - m + 1} \\
 &= \frac{q_m(mq_m + m - 2)}{(mq_m + m - 2)(mq_m + m - 1)} \\
 &= \frac{q_m}{mq_m + m - 1} \quad (A6)
 \end{aligned}$$

式(A5')の右辺と式(A6)の右辺は同じであるから, $t_m(q_m, m - 1) = t_m(q_m + 1, 0)$ である. 式(A4)で $r_m = m - 1$ とすると $A = 0$ となり, 式(A3)で $\Delta t_m(q_m, r_m) = 0$ となるから, この性質は式(A3)でも実際には表現されている. 要するに, 範囲 q_m で余りが最も大きい場合($r_m = m - 1$)とそれよりデータが1つ増えた範囲 $(q_m + 1)$ で余りが無い場合は, 対の数が等しい. 以上をまとめると, 式(A3)で示されるように範囲 q_m で $t_m(q_m, r_m)$ は N とともに単調に増加し, 次の範囲 $(q_m + 1)$ へ移るとき同じ値をとる. したがって, 範囲 q_m における $t_m(q_m, r_m)$ の値は $r_m = 0$ のとき最小で, $r_m = m - 1$ とき最大である. ところで, $m \gg 1$ ならば, 式(A5)と(A6)はそれぞれ次の式(A7)と(A8)のように書き換えることができる.

$$t_m(q_m, 0) = \frac{q_m - 1}{m(q_m - 1/m)}$$

$$\approx \frac{1}{m} \cdot \frac{q_m - 1}{q_m} \quad (A7)$$

$$\begin{aligned}
 t_m(q_m, m - 1) &= \frac{q_m}{m(q_m + 1 - 1/m)} \\
 &\approx \frac{1}{m} \cdot \frac{q_m}{q_m + 1} \quad (A8)
 \end{aligned}$$

式(A7)と(A8)はともに, q_m が十分に大きければ, $t_m(q_m, r_m)$ が $1/m$ に収束する, すなわち,

$$t_m(q_m, r_m) \approx \frac{1}{m} \quad (A9)$$

であることを示している. したがって, m が一定のとき N が大きくなると q_m も段階的ではあるが大きくなるので, N の増加とともに $t_m(q_m, r_m)$ は $1/m$ に収束する. 一方, q_m が小さい場合, 例えば10程度だと, $t_m(q_m, r_m)$ は $1/m$ に較べて約10%小さくなる ($m = 100$ ならば約1%).

ところで, $q_m \gg r_m$ ならば, $q_m \approx N/m$ であるから $m \approx N/q_m$ である. この関係を式(A9)に代入すると,

$$t_m(q_m) \approx \frac{N}{q_m} \quad (A10)$$

なる関係が得られる. 式(A10)は, q_m が一定すなわち N に関係なく1グループを構成するデータの数が一定になるようにグループ分けすると, 計算時間が N に比例して長くなることを示している.

注3: 距離 $\mathbf{h}$ における対の数 $p(\mathbf{h})$ を対の総数 p^Σ で割った値を規格化した対の数 $\phi(\mathbf{h})$ と定義する. すなわち, $p^\Sigma = \sum \mathbf{h} p(\mathbf{h})$ で, $\phi(\mathbf{h}) = p(\mathbf{h})/p^\Sigma$ である. 実験ヴァリオグラム1と2がある場合, 距離 $\mathbf{h}$ における規格化した両者の平均の対の数 $\bar{\phi}(\mathbf{h})$ を, それぞれの規格化した対の数を対の総数を重みとして平均した値と定義すると, 次式が得られる.

$$\begin{aligned}
 \bar{\phi}(\mathbf{h}) &= \{\phi_1(\mathbf{h})p_1^\Sigma + \phi_2(\mathbf{h})p_2^\Sigma\}/(p_1^\Sigma + p_2^\Sigma) \\
 &= \{p_1(\mathbf{h}) + p_2(\mathbf{h})\}/(p_1^\Sigma + p_2^\Sigma)
 \end{aligned}$$

上式より, 実験ヴァリオグラム1と2の距離 $\mathbf{h}$ における平均の対の数 $\bar{p}(\mathbf{h})$ は次式で与えられる.

$$\begin{aligned}
 \bar{p}(\mathbf{h}) &= \bar{\phi}(\mathbf{h})(p_1^\Sigma + p_2^\Sigma) \\
 &= p_1(\mathbf{h}) + p_2(\mathbf{h})
 \end{aligned}$$

なお, $(p_1^\Sigma + p_2^\Sigma)$ は定数であるから, 実験ヴァリオグラム1と2の重み付き相関係数を求めるときの重みは, $\bar{\phi}(\mathbf{h})$ でも $\bar{p}(\mathbf{h})$ としても同じである.

要 旨

対の一部を使うヴァリオグラムの計算における精度の低下と時間の短縮

：正路徹也

地球統計学で用いられるヴァリオグラムを得るには、全データから取り出されるすべての対についてその値の差の二乗を計算しなければならないので、計算時間はデータ数の2乗に比例して長くなる。そこで、通常の統計処理と同様にランダムに抽出した一部のデータを用いてヴァリオグラを計算することが考えられる（一部データ部分計算）。ここでは、それ以外に、データは全部使うが対の相手は先頭部分しか使わない先頭優先部分計算、各データとも近隣のデータとの対しか使わない近隣順部分計算、ほぼ同数のデータからなるグループに分けて、各グループで得られたヴァリオグラムを重ね合わせて全体のヴァリオグラムとするグループ化部分計算法も試みた。部分計算の評価に利用したデータは、小笠原母島付近の8891点の標高に当てはめた平面からの残差である。各部分計算の精度は、部分計算と完全計算で得られたヴァリオグラム間の相関係数で行なった。その結果、一部データ部分計算と先頭優先部分計算の精度は低く、近隣順部分計算とグループ化部分計算の精度は高かった。今回の例では、対の数が完全計算の約1/100である近隣順部分計算とグループ化部分計算の場合、完全計算で得られたヴァリオグラムとの相関係数は0.99以上であった。近隣順部分計算を途中で打切る場合、現段階と1段前の直交ヴァリオグラムの相関係数が 0.9_s (s は小数点以下に並ぶ9の数) ならば、現段階と完全計算の直交ヴァリオグラムの間の相関係数は $0.9_{s/2}$ (9の数が s の1/2程度) が保障される。

キーワード：ヴァリオグラム，計算，部分，時間，標高，精度，効率