

空間サンプリングのシミュレーションによる考察

総務省 政策統括官(統計基準担当) 増成 敬三
早稲田大学経済学研究科 陳 静慧

空間サンプリングは、距離の相関構造を持っている母集団からの抽出と言えて、抽出方法によっては確率変数の独立性を仮定できないため、標本理論の構築が困難な場合もある。例えば、単純ランダムサンプリングを適用すれば、空間従属性の呪縛から解放されて確率変数の独立性が仮定できるため一般的な標本理論が適用できるが、標準誤差が大きくなってしまふ。一方、格子サンプリングは位置情報を利用して抽出地点間の距離が近くなるように抽出するので相関が小さくなり、標準誤差の小さくなるのが期待できる方法であると言える。しかし、空間従属性のために標本理論の構築は困難であり、正方形の領域においてのみ若干の研究がなされていて、判明していることはランダムサンプリングより標準誤差が小さいということだけである。現在のところ正方形の領域でなければ、どのような結果になるのか示されていないのである。

本研究では、正方形ではない領域の空間サンプリングを実施して、どのように標準誤差が変化するかシミュレーションによって明らかにすることを試みることにした。試す方法は、1. 単純ランダムサンプリング、2. 格子サンプリング、3. 格子ランダムサンプリング、4. ラテン格子サンプリングである。

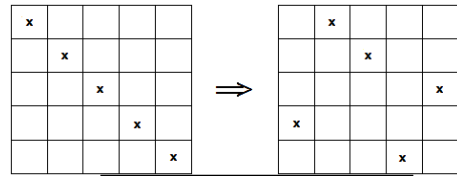


図1 ラテン格子サンプリング

格子ランダムサンプリングは、抽出地点を格子サンプリングのセル内位置をランダムに選択する方法であり、また、ラテン格子サンプリングは、各行と各列に1点ずつのみ抽出点を配置する方法で、位置決めの方法は図1の左のように対角セルを選択しておいて、行または列をランダムに入れ替えて抽出セルを決める方法である。

シミュレーションの対象として図2のようなランダム曲面を生成した。これは0から π までのCosをz軸を中心に360度回転させたコーン型を生成し、その位置、高さ、半径をランダムに変化させて平面上に配置して曲面を生成したものである。それを図3で示すようなランダムな領域形状を生成したものと論理積を計算して図4に示すような、値と領域形状がランダムな曲面を生成して、標準化して利用した。

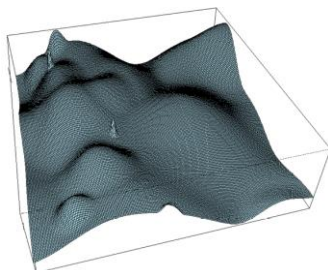


図2 ランダム曲面

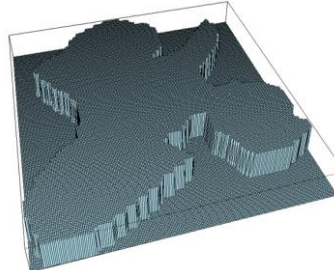


図3 ランダム領域

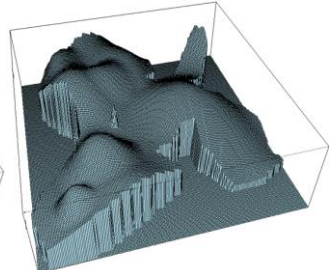


図4 ランダム曲面領域

シミュレーションは標本サイズを変えながら5000回ずつ実施した。結果

を図5に示す。どの方法でも標本サイズが大きくなるにつれて標準誤差は小さくなり、試した方法のなかでは単純な格子サンプリングが最も標準誤差が小さくなった。格子サンプリングの場合、領域内に含まれる抽出点の数は一定にならず多少のばらつきがある。よって、同程度の標本サイズになった結果を比較できるように集計した。

格子サンプリングには多くの発展形が存在する。格子セルのいくつかを選択する方法や、部分的な格子セルの面積を変える方法、また同面積に分割した領域から同数セルを選択するなど、さまざまである。それらの方法を組み合わせてアンサンブル平均(位相平均)する方式なども提案されている。

その他、いくつかのシミュレーション結果も得ているので、詳しい結果については当日報告する。

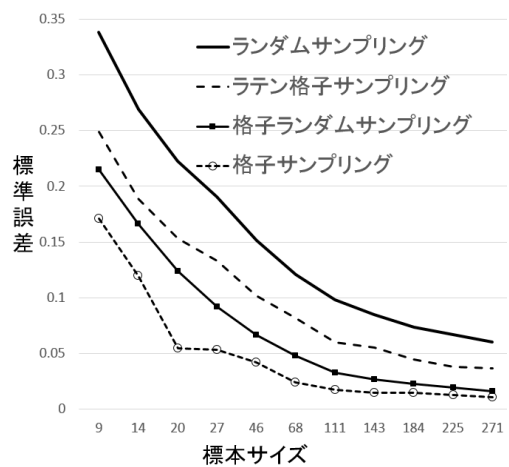


図5 方法毎の標準誤差