

統計解析に適した匿名加工化法

(株)ジャパン・ティッシュ・エンジニアリング 松木 伯元

医療情報などの個人情報を第三者が取り扱う際は、データに匿名加工を施すことが望ましい[1]。しかし、通常は加工されたデータをそのまま解析することができない。

ところで多くのパラメトリックな統計手法（Student の t 検定や Pearson の相関係数や回帰分析など）は、計算の観点からは内積と平均の計算に帰着される。そのため内積と平均を保存する変換によってデータを変換したとき、変換前後でこれらの統計手法による結果は不変である。カーネル法[2]はこの著しい性質（ここでは便宜的に不変性と呼ぶことにする）を非線型データ解析に利用している。ここで第三者に個人情報の統計解析を依頼する状況を想定しよう。上記の不変性に加えて、第三者が変換データから元データを復元することが不可能であるなら、変換データを解析することで、元データを保護しつつ目的とする結果を得ることができる。

本講演では、この不変性を利用した匿名加工法の具体案を提示する。最初に、乱数を用いて各列の成分の和が 1 になる直交行列を生成する簡単な方法[3]について述べる。この直交行列によるデータ変換のもとで、データ・ベクトルの内積と平均が不変であることは明らかである。さらに、サンプル数に変数の数よりも多いとき、直交行列の情報なしに変換データから元データを復元することはほぼ不可能である。最後に、この直交行列によるデータ変換によって種々の統計手法の結果が不変であることを数値例とともに示す。

参考文献

- [1] 宇賀克也, 次世代医療基盤法の逐条解説, 有斐閣, 2019.
- [2] 福水健次, カーネル法入門 – 正定値カーネルによるデータ解析 –, 朝倉書店, 2010.
- [3] N. Matsuki, A construction of special orthogonal matrices, submitted for publication.