

方向データの相関

大阪府立大学 綿森 葉子

1 はじめに

方向データは円周上のデータとみなすことができ、風向や渡り鳥の飛び立つ方角などが代表的である。一般的には、球面やさらに次元の高い超球面上のデータも含むが、実際の応用例はそれほど多くない。また、円周上のデータの解析手法を球面上に拡張するのは、通常のユークリッド空間に比べると簡単ではない場合が多い。

一般次元の超球面上のデータ対について、これまでに、いくつかの相関係数が定義されている。これは、ピアソンの相関係数のような決定的な相関係数が円周上では定義できていないことを意味する。では、相関とはなんであろうか？まずはその意味を再確認してみる必要がある。参考文献として本を数冊あげておく。([1] ~ [3])

2 相関

極座標を用いて \mathbb{R}^p の特異な分布として捉えた場合、相関係数はそのまま定義できるが、ノルム制約が大きく影響するため回帰関数はそうはいかない。では、角度をもとにしてはどうか？今度は実数倍が自然に定義されないため、新しい概念を導入する必要に迫られる。

そもそも相関とはどのようなものかと考えるのが適当だろうか？1つの答えとして、2つの変数間の依存性を測るものと捉えることにする。方向データ間の相関を考えると、(i) 0方向の選び方によらない(回転によって不変)、(ii) 依存性が強ければある最大値(例えば1)に近づき、弱ければある最小値(例えば0)に近づく、という性質を満たしていることを条件とする。

方向データはふつう、16方位、360度までの角度、 $[0,1]$ 区間など値としては限られた範囲で得られる。適切な順位を定めることさえできれば、順位に基づくノンパラメトリック法を用いることは1つの手段となりうる。ここでは、 \mathbb{R}^1 上に射影したデータから得られる順位をもとにスピアマンの順位相関を求め、全ての回転の中で最大値を相関と定義する。上記(i),(ii)の性質を満たすことは明らかである。ただし、最大値を与える回転は1つとは限らない。

参考文献

- [1] N.I.Fisher, Statistical analysis of circular data, 1993 Cambridge University.
- [2] S.R.Jammalamadaka and A.SenGupta, Topics in Circular Statistics, 2001 World Scientific.
- [3] K.V.Mardia and P.E.Jupp, Directional Statistics, 2000 John Wiley.