

起点・終点点パターンの統計モデリング

明治大学 商学部 城田慎一郎

近年、社会科学・環境科学など多様な分野で空間点パターンの統計解析が注目を集めている。空間点パターンとは、犯罪の発生場所や植物の植生パターンなど、空間的な位置情報を伴うデータの総称である。本研究では、このような空間点パターンの中でも、起点と終点がペアで得られている点パターンに対する統計モデリングを紹介する。このような点パターンは、車両の盗難場所(起点)と発見場所(終点)、自宅の住所(起点)と勤務先の住所(終点)など、実社会の様々なシーンにおいて見られる。また、車両の盗難場所と発見場所といったデータにおいては、発見場所が欠損する(盗難された車が見つからない)といったケースも考えられ、終点に欠損がある場合も考慮する必要がある。本研究の実データ分析においては、メキシコとブラジルの都市における車両の盗難場所と発見場所の点パターンデータに対し、発見場所の欠損がある場合(メキシコ)とない場合(ブラジル)のそれぞれのケースに対して、別々の方向性に基づく統計モデリングを提案している。

空間点パターンに対する統計モデリングとしては、空間点過程が用いられる。空間点過程の代表的なクラスとしては、ポワソン過程があり、空間上に強度関数を定義し、場所に依存する点の発生確率を導入することができる。より発展的なモデルとしては、対数ガウスコックス過程(log Gaussian Cox processes, LGCP)と呼ばれる確率過程のクラスがあり、強度関数の対数がガウス過程に従う。ガウス過程は、空間統計において、空間相関を柔軟に導入できる重要な確率過程であるが、そのサンプリングにおける計算コストが高いことが知られている。しかし、近年のガウス過程に対する高速かつ精度の高い近似手法の発展に伴い、今後ますますLGCPの利用が促進されると考えられる。本研究でも、LGCPをベースに、起点・終点のペアに対する統計モデリングを構築している。実データでは、終点は起点からの距離に依存して、その存在確率が変動しやすい(起点に近い場所に終点が存在する)という事象が観測される傾向があるため、本モデルでも起点に依存するカーネル関数を導入し、起点と終点との依存構造もモデルに取り込んでいる。実データの分析を通じて、このモデルの有用性を検証している。