

自殺予防に向けた地域統計の時空間解析と要因分析

■ 自殺死亡の地域統計

「自殺」という言葉を聞いたときどのようなイメージを抱きますか? 「日本は自殺者が多い」、「14年連続で3万人を超えた」、「人身事故で電車が止まった」、「ゲートキーパー」などさまざまでしょう。残念ながら日本は世界の各国の中で最も自殺率が高い国の一つです。この問題に対処すべく、我々はまず、国立精神・神経医療研究センター精神保健研究所からの受託研究のもと「自殺対策のための自殺死亡の地域統計」を作成しました。このデータは地域ごとの、時間、年齢、性別、自殺の手段、配偶関係、職業などの観点から自殺死亡を集計したデータです。

これらの集計されたデータの特徴をとらえるために、見える化にも取り組んでいます。例えば、コロプレスマップを用いた自殺率の塗り分けや、Generalized Association Plots(GAP)を用いた時空間データの行列表現の可視化や、RnavGraphを用いて年齢階級別のデータをクラスター分析した結果の可視化などです(図1)。

■ 時空間集積性の検出

地域統計を利用している利点を生かし、まず空間集積性の検出、つまりどここの地域に自殺者が多いのかを探ることから始めました。図2は検出された集積エリアとそれを求めるために用いた走査法のためのエシェロンデンドログラムです。さらに、時間的な変遷を考慮して時空間的な集積性、関東地方に着目しての解析、GISを利用した地域の大きさの変更、時間的な増減を考慮した集積性、希少地域の集積性などについても検出を試みています。

■ 自殺の要因分析

先述の自殺死亡の地域統計と同じ時期・場所のさまざまな変数をつなぎ合わせ、自殺の理由を探るべく様々な解析を行っています。その1つが、最も関係があると思われる「婚姻関係」(特に配偶者との死別)、「職業」(特に完全失業者)および自殺率の空間的自己相関を考慮した空間相関分析です。また、「気象」や「地理」に関連

する変数を用いた回帰分析や、そして経済的な要因を加えた因果分析、共分散構造分析についても取り組み始めています。さらには、自殺の背景的な要因を考え、メンタルヘルスに関わる専門家との共同研究を通じて医療的な要因に関連する項目についても取り組んでまいります。

■ 研究集会とこれからの展望

これらの解析の結果を様々な機会に報告するとともに、様々な分野の関係者と共に研究集会を通じて議論し、効果的な健康保健政策の提唱につなげることを目指します。2012年度は10月頃に研究集会の開催を予定していますので、ご関心いただきましたら是非参加して議論いただければ幸いです。

久保田 貴文

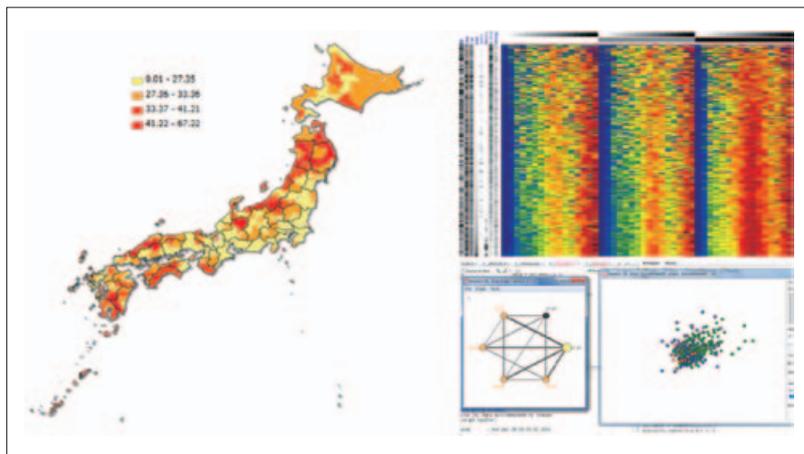


図1: 1988年~1992年の自殺率のコロプレスマップ(左)とGAPを用いた行列表現(右上)、RnavGraphを用いたクラスター分析の結果(右下)

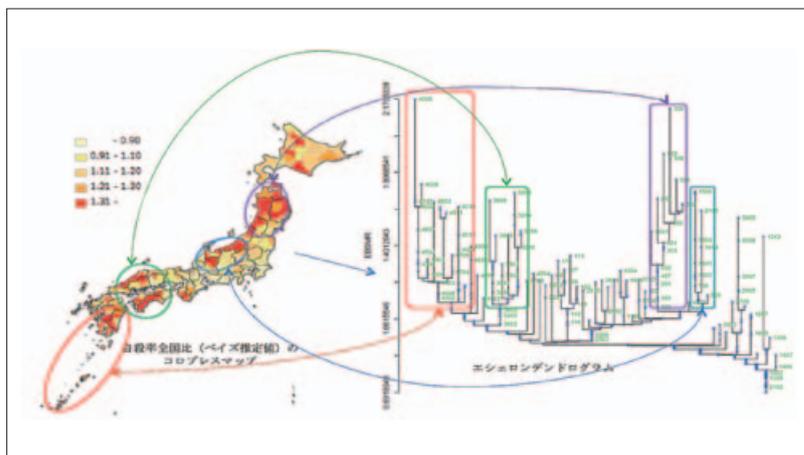


図2: 1988年~1992年の自殺率全国比(ベイズ推定値)のコロプレスマップおよび集積エリア(左)とエシェロンデンドログラム(右)との対応

ゲノム解析プロジェクト

■ プロジェクトのミッション

当プロジェクトは2012年1月に発足したばかりの新しいプロジェクトです。個別化医療や生物多様性保全など現代社会が抱える様々な課題を解決するために有用な知見を大量のゲノム情報から抽出するための方法論と、発見された知見を活用して社会に潜在するリスクを定量化するための手法の開発を通じて、ゲノムを基盤としたリスク情報の理解に貢献することを目指しています。疾患、薬剤感受性や雑種強勢などに関わる遺伝子をゲノムから同定するにあたり必要になる多重性調整や影響分析などの方法論の整備、ゲノム情報による生物多様性や病原性進化の解析、アミノ酸の構造解析、ゲノム情報の漏洩リスクの定量化などを扱っています。以下では、プロジェクトの研究の一つを紹介いたします。

■ 遺伝子発現差解析におけるP値推定の問題点

遺伝子発現差解析は遺伝子発現データに基づいて頻繁に行われています。実際には、適当な検定統計量に基づいて、対応するP値が推定されて、その値で判断されることが多いです。そのP値は並べ替え標本などに基づいて推定されています。しかし、遺伝子発現データは、実験回数がしばしば非常に小さく、その結果として十分な数の並べ替え標本が得られないことがあります。そういうときはP値の推定精度に不安が生じます。

そこで、ある遺伝子におけるP値を推定するときに、他の遺伝子のデータを援用することで、並べ替え標本の数を増やす試みが行われています。遺伝子発現データで

は、実験回数は小さいけれども、遺伝子数は非常に大きい（たとえば数千となる）ことが一般的だからです。しかし、P値は帰無仮説が正しい下で定義されているのに、遺伝子の中には帰無仮説が正しいものと正しくないものが混在しているので、このままではP値の推定は妥当ではありません。

この問題を克服するために、他の遺伝子のデータを適切に援用できるように検定統計量をうまく構成して、対応する帰無統計量に基づいてP値を妥当に推定するというアイデアが提案されました。このアイデアは興味深いのですが、検定統計量の構成がアドホックなので、検出力がより高い方法があるのではないかと疑問は残り続けていました。

■ 数理的思考による新しい手法の発見

本研究では、他の遺伝子のデータを適切に援用するための条件を、アドホックにではなく数理的にきちんと整理して、そのような条件の下で最も検出力が高い検定方法を導出しました。根幹となる条件を数理的にうまく整理することで、アドホックさだけでは発見しにくい優れた方法を導出することに成功しました。そこで得られた検定はアドホックに提案された検定よりも検出力が高いことが証明されます。さらに、検出力の差は現実的にも非常に大きく、現実のデータ解析でも興味深い結果を与えており、インパクトのある結果が得られました。

間野 修平、藤澤 洋徳

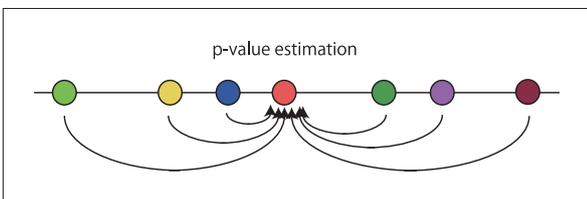


図1: 他の遺伝子のデータを援用したP値推定

$$\hat{p}(T(Z^*); t_g) = \frac{1}{B} \sum_{b=1}^B I(T(Z_{bg}^*) > t_g)$$

$$\Downarrow$$

$$\hat{p}(T(Z^*); t_g) = \frac{1}{BG} \sum_{b=1}^B \sum_{g'=1}^G I(T^{\text{null}}(Z_{bg'}^*) > t_g)$$

図2: P値の推定式。上の式は標準的なP値の推定式。下の式は他の遺伝子データを援用したP値の推定式で右辺に工夫が施されている。

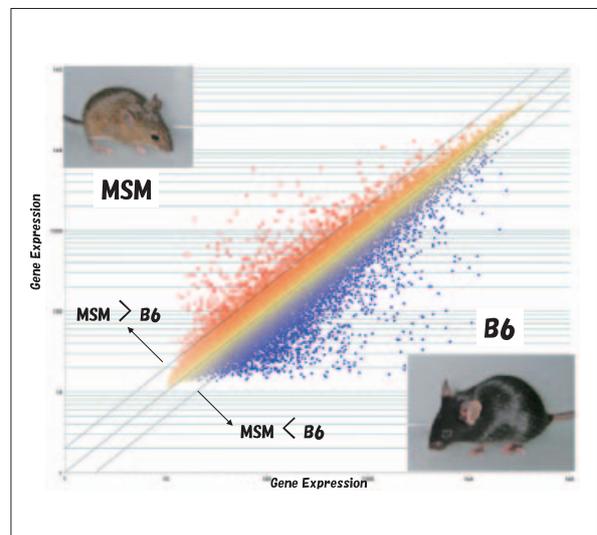


図3: マウスのB6とMSMの遺伝子発現差 (提供: 国立遺伝学研究所哺乳動物遺伝研究室)