

<修士論文>

新型コロナウイルス感染症に
関する考察
(要旨)

滋賀大学大学院
データサイエンス研究科
データサイエンス専攻

修了年度 : 2020年度
学籍番号 : 6019121
氏名 : 横内 淳史
指導教員 : 姫野 哲人
提出年月日 : 2021年1月20日

2019年に発生した新型コロナウイルス感染症（国際正式名称：COVID-19）が引き起こした世界的流行（パンデミック）によって、世界中の人々が日々の生活に大きな影響を受けた。日本においても同様で、2020年11月現在でも感染収束には至っておらず、第3波とよばれる感染拡大が発生している。これに伴い、感染拡大抑制について様々な対策が実施されている。この取り組みにはデータに基づいた対策の立案が非常に重要となるが、入手できる情報に欠損やバラつきも多い。そのような制約下でもデータサイエンスを活用し、どのようなことが明らかにできるか検討した。

本研究では2つのテーマについて検討した。

1つ目は、感染拡大のごく初期段階における感染状況の把握である。2020年11月時点では、陽性者数だけでなく、PCR検査実施人数や、退院又は療養解除となった者の数、死亡者数などが政府からオープンデータとして公開されている。しかし、感染拡大の初期段階では、入手できる正確な情報は少なかった。その限られた情報の中でも、感染拡大の状況を正しく把握することは重要な問題である。本研究では、日々の新規陽性者数から算出される累計陽性者数のグラフに着目した。このグラフを活用して感染状況を可視化することで、一般の人々が理解しやすくなるのではないかと考えた。本研究では、感染の状況を新規陽性者数のグラフに重ね合わせて可視化する。

2つ目は、潜在的感染者の推定である。感染症の流行を記述するモデルのひとつとして、SIRモデルが知られている。しかし、SIRモデルは決定論的に記述するモデルであるため、単純なSIRモデルでは新型コロナウイルス感染症の流行をモデル化することは難しい。最も重要な問題は、感染率が時間とともに変化することである。本研究では、パラメータの時間変化をモデルに取り込んだ。また、実効再生産数は、1人が平均して何人に感染させるかを示す指標である。西浦（2020）は、週ごとの新規感染者数を用いた近似的な実効再生産数の推定を提案している。本論文ではこの実効再生産数を活用し、新型コロナウイルス感染症の状況をSIRモデルをベースとしてモデル化し、潜在的感染者を推定した。潜在的感染者数は、感染対策を検討するうえで非常に重要な指標であり、有効な対策を決定するのに役立つはずである。

本研究では、2020年5月31日までを第1波、2020年6月1日から2020年9月30日までを第2波、2020年10月1日以降を第3波と呼ぶこととする。感染状況の把握については、厚生労働省がオープンデータとして公開している陽性者数のデータを活用し、初めて陽性が判明した2020年1月16日から2020年5月31日までを対象とした。本研究では、日々の累計陽性者数をロジスティック回帰分析を行い、各時点の累計陽性者数に最小二乗法によるパラメータの推定を試みた。推定したパラメータは最終陽性者数と感染率で、感染収束日は最終陽性者数の99%に到達する日付とした。

また、最小二乗法により推定したロジスティック曲線の信頼区間について考えた。各パラメータが漸近正規性を持てば、推定値 $\pm 1.96 \times$ 標準偏差を計算することで信頼区間が得られるが、一般には漸近正規性を持たない。したがって、推定値 $\pm 3 \times$ 標準偏差を信頼区間とし、直近のデータが信頼区間を外れた場合、感染状況に変化があったと見なす。

潜在的感染者の推定については、まず、日々の新規陽性者について検討した。日本ではPCR検査で陽性となった者は、基本的に病院やホテルに隔離されるため、SIRモデルにおける隔離者 $R(t)$ ととらえるべきである。これに対し感染者 $I(t)$ は、まだPCR検査で陽性となっていない潜在的感染者ととらえることができる。これを状態空間モデルを用いた階層ベイズ推定法で推定した。潜在的感染者 $I(t)$ を推定することで、どのくらい感染が広がっていたかを推定できる。

実データを用いた感染拡大状況の把握については、各時点の直近1週間の陽性者数の推移をそれ以前の感染状況から変化がないか、分析を行った。その結果、2020年3月10日、2020年3月25日、2020年4月19日で変化が検出できた。2020年3月10日および、2020年3月25日については感染拡

大傾向であるが、2020年4月19日は感染縮小傾向で、緊急事態宣言により感染収束に一定の効果があつたことを示している。また、累計陽性者数から各時点の新規陽性者数を算出し、感染拡大予測を可視化した。2020年4月12日時点での最大感染者数の推定値が約2,755人に対し、2020年5月31日時点での最大感染者数の推定値は約464人で、感染拡大を約83%抑制することができたと考えられる。また、感染収束日は2020年6月28日から2020年6月3日と25日短縮できたと考えられる。

また、状態空間モデルから、第1波では陽性率の高さの影響もあり検知率が低いと推定された。第2波、第3波と比較しても相対的に検知率が低く、第1波の新規陽性者のピーク付近の2020年4月13日の検知率は約17.2%となった。それに伴い潜在的感染者数も約3,350人となり、第1波の新規陽性者のピークである708人の4.73倍となった。これに対し、第2波では2020年8月4日に検知率約25.2%、潜在的感染者数約5,341人となった。これは第2波の新規陽性者のピークである1595人に対して3.35倍であった。また、新規陽性者が少なかった2020年5月26日付近では、潜在的感染者数に対して十分なPCR検査が実施され、2020年5月23日の検知率は約88.2%と高かった。さらに、2020年9月29日前後にPCR検査数が非常に多く、陽性率が低い期間では、2020年5月26日付近と同様に、2020年9月29日に検知率が約81.7%と高く推定された。しかし、第2波・第3波間の新規陽性者が高止まりとなったその他の期間では40~50%程度と低くなった。本モデルで推定された2020年10月1日までの、潜在的感染者数を含む、累計推定感染者数は約94,537人で、累計陽性者数と比較すると1.13倍となった。また、日本の人口の約125,710,000人を踏まえると、約0.08%が感染したと推定される。