

<修士論文>

新型コロナウイルス感染症に
関する考察

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

目次

1	はじめに	2
1.1	研究の背景	2
1.2	研究の目的	3
2	提案手法	5
2.1	感染拡大状況の把握	5
2.2	潜在的感染者の推定	7
3	検証結果	11
3.1	感染拡大状況の把握	11
3.2	潜在的感染者の推定	17
4	まとめと今後の課題	19

新型コロナウイルス感染症に関する考察

滋賀大学データサイエンス研究科 横内 淳史

要 旨

日本における新型コロナウイルス感染症の感染者数について、各時点においてロジスティック回帰分析を行い、それから推定される信頼区間に基づき感染状況の変化について考察を行った。その結果、2020年5月31日までの第1波の流行では、2020年3月10日と2020年3月25日、2020年4月19日に感染状況の変化をキャッチアップすることができることを示した。また、2020年4月7日に発令された緊急事態宣言によって、累計感染者数を約83%抑制する効果があったことを示した。

さらに、日々の新規感染者数およびPCR検査数を用いて状態空間モデルを構築し、潜在的感染者数を階層ベイズ推定法で推定した。その結果、2020年9月30日までの潜在的感染者が、最大で5,000人以上と存在したと推定された。また、累計推定感染者数は約9,400人となり、累計陽性者数の約1.13倍と推定されることを示した。

1 はじめに

1.1 研究の背景

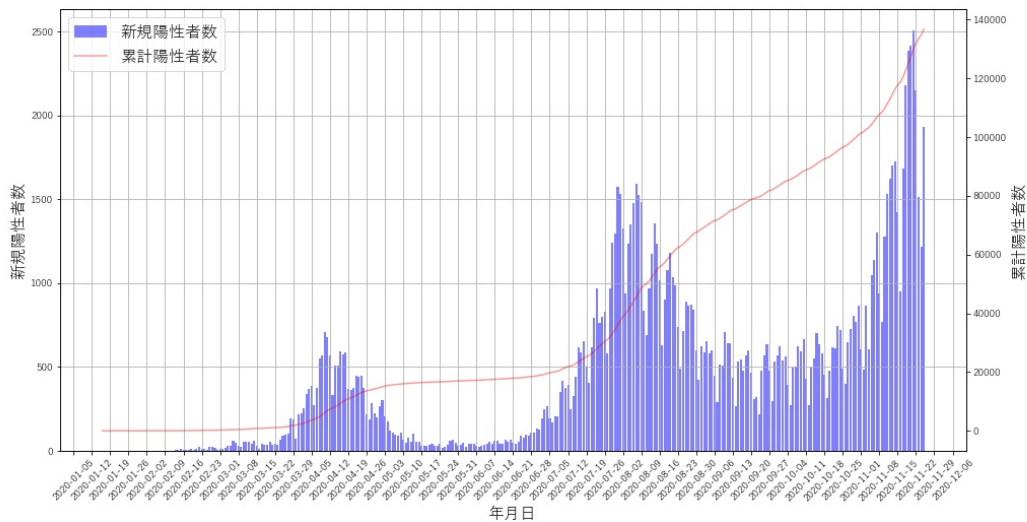


図 1: 日本における新型コロナウイルス感染症の感染状況

2019年に発生した新型ウイルス感染症（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に、図1に示すとおり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を検討した。

1.2 研究の目的

本研究では2つのテーマについて検討している。

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 [1]。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。このため、中野・池田 (2020) は、感染拡大率を表す指標である K 値を提案している [2]。 K 値は週あたりの感染者数の増加率を表し、次の式で表される。

$$K = 1 - \frac{1 \text{ 週間前の総感染者数}}{\text{総感染者数}}$$

この K 値は一定の割合で減少すると仮定し、その直線と日々の K 値の乖離から感染状況を可視化することを提案している。これに対し、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。新規陽性者数や累計陽性者数については連日マスコミでも報道されており、新規陽性者数のグラフなどは一般の人々も目にすることが多い。このグラフを活用して感染状況を可視化するほうが、一般の人々が理解しやすい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。新型コロナウイルス感染症のような感染症の流行においては、国民一人一人に状況を認識してもらうことも重要であり、それに役立つはずである。

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 [3]。 S (susceptible) は感染症に対する免疫を持たず感染する可能性がある人数を表し、 I (infected) は感染している人数を、 R (recovered) は感染症から回復し免疫を獲得した人数もしくは亡くなった人数を表す。これらの関係は次の微分方程式で記述される。

$$\frac{dS}{dt}(t) = -\beta S(t)I(t) \quad (1)$$

$$\frac{dI}{dt}(t) = \beta S(t)I(t) - \gamma I(t) \quad (2)$$

$$\frac{dR}{dt}(t) = \gamma I(t) \quad (3)$$

β は感染率、 γ は回復率を表す。また、この微分方程式から、感染期間において1人の感染者が何人に感染させるかを表す指標、基本再生産数 \mathfrak{R}_0 は以下のように算出される [4]。

$$\mathfrak{R}_0 = \frac{\beta S(0)}{\gamma}$$

基本再生産数はその感染症の感染力の強さを表す指標として用いられるが、新型コロナウイルス感染症においては、まだ値が確定していない。これに対し、実効再生産数 $\mathfrak{R}(t)$ と呼ばれる指標が注目

されており、次のように定義される。

$$\mathfrak{R}(t) = \frac{\beta S(t)}{\gamma} \quad (4)$$

実効再生産数は各時点の感染状況を表し、 $\mathfrak{R}(t) > 1$ であれば感染は拡大し、 $\mathfrak{R}(t) < 1$ であれば感染は収束に向かっていると判断できる。ただし、回復者を表す $R(t)$ と実効再生産数を表す $\mathfrak{R}(t)$ を混同しないよう注意されたい。

ここで、SIR モデルは決定論的に記述するモデルである。したがって、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 β が時間とともに変化することである。例えば、感染拡大に伴う外出自粛要請やロックダウン、マスクや手洗いの浸透などにより感染率は変化する。そもそもウイルスが変異していることも報告されており、それに伴う感染力の変化も考えられる。このようなパラメータの時間変化をモデルに取り込む必要がある。

また、実効再生産数 $\mathfrak{R}(t)$ も同様に感染率 β の時間的変化の影響を受ける。実際の感染症における実効再生産数の推定は非常に重要であり、新型コロナウイルス感染症においても経済活動の再開なども実効再生産数の推移を確認する必要があるとされている [5]。さらに専門的な知見がなくても実効再生産数を推定できるよう、Excel を活用した推定方法なども提案されている [6]。日本における新型コロナウイルス感染症に対しても実効再生産数の推定は試みられており、政府の政策決定にも活用されている。

そんななか、西浦 (2020) は発症日に着目した実効再生産数の推定について、問題点の議論および対応案の提案を行っている [7]。発症日に基づく実効再生産数の推定では、感染発覚時に調査を行う必要があり、そもそもプライバシーの問題から公開されるとは限らない。リアルタイム性に課題があるため、週次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。この推定法を用いた実効再生産数については、東洋経済オンラインでも公開されており [8]、ここでは次のような近似式が用いられている。

$$\mathfrak{R}(t) = \left\{ \frac{\text{直近 7 日間の新規陽性者数}}{\text{その前 7 日間の新規陽性者数}} \right\}^{\text{平均世代時間/報告間隔}} \quad (5)$$

ここで、平均世代時間は 5 日、報告間隔は 7 日と設定されてる。報告間隔は直近およびその前の 7 日間の新規陽性者数を用いているためであり、平均世代時間は、感染源の感染から 2 次感染者の感染までに要する時間を西浦の研究結果から 5 日と仮定している。

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モデル化した。式 (3) および式 (4) の感染者数および回復者と式 (5) の実効再生産数から、次のモデル式で記述することができる。

$$I(t) = \gamma(t-1)\mathfrak{R}(t-1)I(t-1) - \gamma(t-1)I(t-1) + I(t-1) \quad (6)$$

$$R(t) = \gamma(t-1)I(t-1) + R(t-1) \quad (7)$$

その時点の感染者 $I(t)$ は、前日の感染者 $I(t-1)$ が 1 人当たり回復率 $\gamma(t-1)$ と実効再生産数 $\mathfrak{R}(t)$ の積の人に感染させ、前日の感染者は回復率 $\gamma(t-1)$ で回復する。このように、実効再生産数を活用することで未感染者 $S(t)$ を用いずモデル化でき、このモデルをベースとすることでパラメータ推定が容易となる。

本研究では、このモデルをもとに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。

2 提案手法

2.1 感染拡大状況の把握

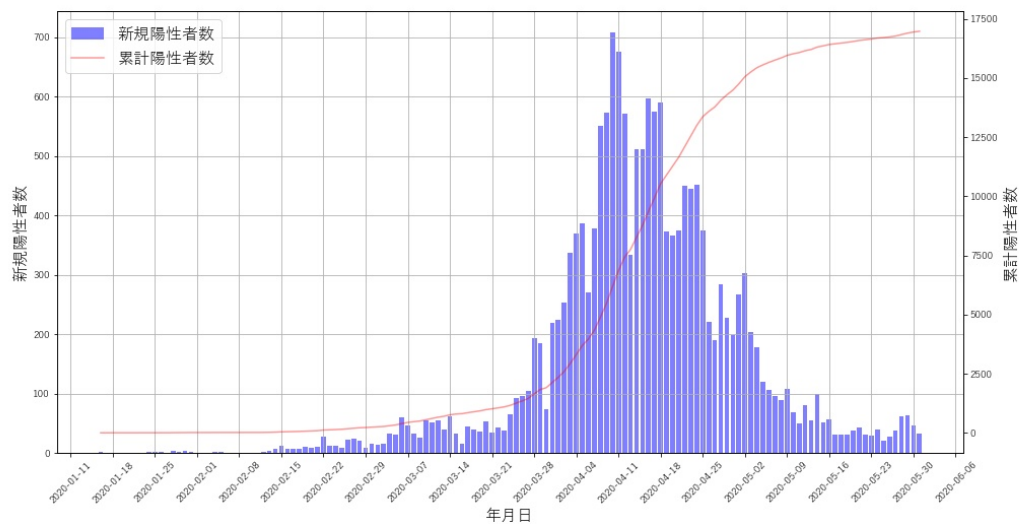


図 2: 第 1 波における感染状況

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。まず、第 1 波の感染状況を図 2 に示す。データ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 日までを対象としている。

棒グラフが新規陽性者数を表し、曲線が累計陽性者数を表す。特に累計陽性者数のグラフに注目すると、見事なシグモイドカーブを描いている。これは日本だけの特徴ではなく他の国でも同様で、最初に感染が広がった中国でも同様の特徴を示す。このことから、日々の累計陽性者数をロジスティック回帰分析することで、どの程度感染が広がるか推定できるのではないかと考えた。

ロジスティック方程式は、次の微分方程式 (式 (8)) で記述される。

$$\frac{dN}{dt} = rN \left(1 - \frac{N}{K} \right) \quad (8)$$

N : 累計陽性者数, K : 最終陽性者数, r : 感染率

この微分方程式を解くことで、各時点の累計感染者数 $N(t)$ は次の式 (9) で記述できる。

$$N(t; K, r) = \frac{K}{1 + (K/N_0 - 1)e^{-rt}} \quad (9)$$

N : 累計陽性者数, N_0 : 初期陽性者数, K : 最終陽性者数, r : 感染率

ちなみに、これらは先に挙げた SIR モデル (式 (2)~(4)) における回復率 γ を 0 としたモデルと同義ととらえることができる。この式 (9) について各時点の累計陽性者数のデータを用い、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 K と感染率 r で、感染収束日を最終陽性者数の 99% に到達する日付とした。

分析着手時点では、最終陽性者数や収束期間が早期に推定できるのではないかと期待していた。しかし、時間を進めるごとに、最終陽性者数 K および感染率 r は大きく変動した。これを逆に、感染状況のステージ変化をとらえることができるのではないかと考えた。

ここで、推定値の信頼区間について考える。データ $(t_i, N_i) (i = 0, 1, \dots)$ を用いて最小二乗法により推定した K と r をそれぞれ \hat{K} , \hat{r} とする。このとき、これらの推定量は漸近不偏性を持つ。つまり、 $\lim_{n \rightarrow \infty} E[\hat{K}] = \dot{K}, \lim_{n \rightarrow \infty} E[\hat{r}] = \dot{r}$, \dot{K}, \dot{r} は真の値とする。さらに、 $N(t; \hat{K}, \hat{r})$ をテイラー展開することにより、

$$N(t; \hat{K}, \hat{r}) - N(t; \dot{K}, \dot{r}) \approx N_K(t; \dot{K}, \dot{r})(\hat{K} - \dot{K}) + N_r(t; \dot{K}, \dot{r})(\hat{r} - \dot{r})$$

が得られる。ここで、

$$N_K(t; \dot{K}, \dot{r}) = \frac{\partial}{\partial K} N(t; \dot{K}, \dot{r}) = \frac{1 - e^{-rt}}{(1 + (\dot{K}/N_0 - 1)e^{-rt})^2}$$

$$N_r(t; \dot{K}, \dot{r}) = \frac{\partial}{\partial r} N(t; \dot{K}, \dot{r}) = \frac{t\dot{K}(\dot{K}/N_0 - 1)e^{-rt}}{(1 + (\dot{K}/N_0 - 1)e^{-rt})^2}$$

である。これらの結果より、

$$\lim_{n \rightarrow \infty} E[N(t; \hat{K}, \hat{r})] = N(t; \dot{K}, \dot{r})$$

$$\lim_{n \rightarrow \infty} V(N(t; \hat{K}, \hat{r})) = N_K(t; \dot{K}, \dot{r})^2 V(\hat{K}) + N_r(t; \dot{K}, \dot{r})^2 V(\hat{r})$$

$$+ 2N_K(t; \dot{K}, \dot{r})N_r(t; \dot{K}, \dot{r})Cov(\hat{K}, \hat{r})$$

であり、求めたい曲線について各点 t で漸近有効性を持ち、それぞれの点の漸近分散（および漸近標準偏差）も計算できる。しかし、 \dot{K}, \dot{r} は未知なので、推定量として置き換え、式 (10) を得る。

$$\lim_{n \rightarrow \infty} V(N(t; \hat{K}, \hat{r})) = N_K(t; \hat{K}, \hat{r})^2 V(\hat{K}) + N_r(t; \hat{K}, \hat{r})^2 V(\hat{r})$$

$$+ 2N_K(t; \hat{K}, \hat{r})N_r(t; \hat{K}, \hat{r})Cov(\hat{K}, \hat{r}) \quad (10)$$

\hat{K}, \hat{r} 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 \hat{K}, \hat{r} は漸近正規性を持たない。しかし、推定値 $\pm 3 \times$ 標準偏差 とすれば分布を問わず信頼区間が真の値を約 90% の確率で含むので、推定値 $\pm 3 \times$ 標準偏差 を信頼区間とした。

感染拡大状況の分析アルゴリズムを下記に示す。

1. 各時点の 1 週間前までの累計陽性者数を用いてロジスティック回帰分析を行う

2. 式 (10) に基づき標準偏差を算出, 推定値 $\pm 3 \times$ 標準偏差 を信頼区間とする

3. 直近 1 週間の累計陽性者数をプロットする

このアルゴリズムにより, 2020 年 3 月 6 日時点のデータを例に分析した結果を図 3 に示す. この例では, 直近 7 日間のデータが信頼区間を外れていないが, 信頼区間を外れた場合, 前週までの感染状況から変化があったと見なす.

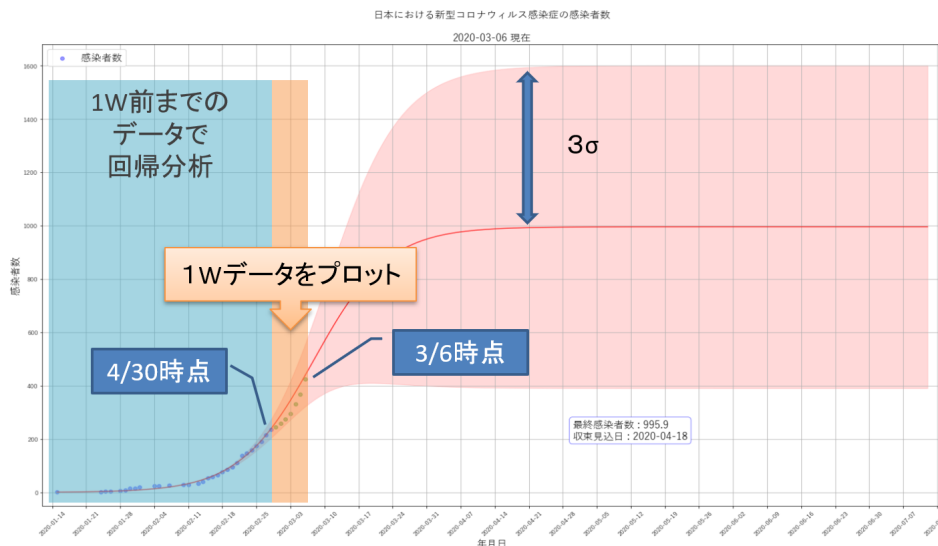


図 3: 感染拡大状況の分析方法

2.2 潜在的感染者の推定

まず, 日々の新規陽性者について考える. 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. このため, 他者に感染させるリスクは大幅に下がるはずである. SIR モデルにおける R は隔離者とも考えることができ, 日々の累計陽性者は感染者 $I(t)$ ではなく, 隔離者 $R(t)$ ととらえるべきである. これに対し感染者 I 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. 状態空間モデルは, 観測できない「状態」を観測データから推定することができる. まず, 観測データを新規陽性者とした. 新規陽性者は $R(t)$ の日々の増加分であり, 式 (7) の右辺の第一項を意味する. この増加分を $\Delta R(t)$ とし, 次の式で定義する.

$$\Delta R(t) = \gamma(t-1)I(t-1) \quad (11)$$

さらに, 未観測データを潜在的感染者 $I(t)$ とし, $I(t)$ 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. モデルを構成する基礎的な数値について, 図 4, 5 に示す. データは 2.1 節と同様に, 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および, PCR 検査実施件数をもとにしている.

まず, 図 4 は, 式 (5) にて算出される実効再生産数 $\mathfrak{R}(t)$ および新規陽性者数 $\Delta R(t)$ の推移を表す. ここで, 実効再生産数 $\mathfrak{R}(t)$ は 7 日間の新規陽性者数をもとに算出しているため, 新規陽性者数

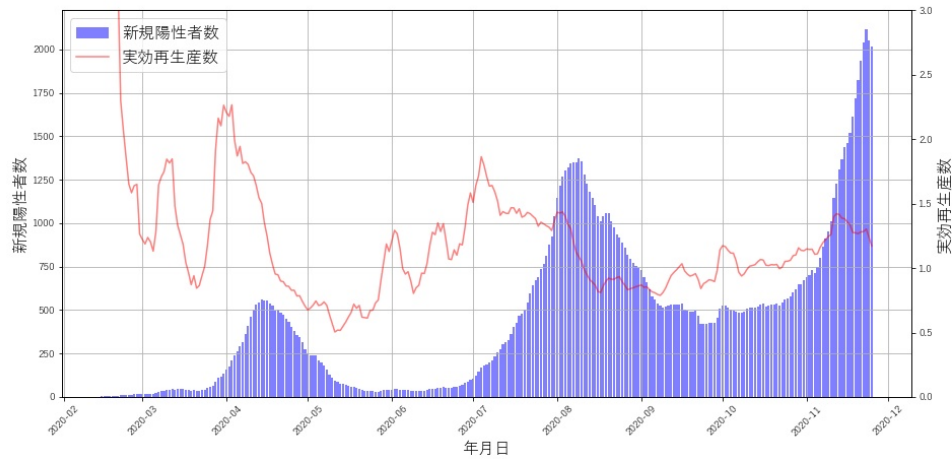


図 4: 新規陽性者数および実効再生産数

$\Delta R(t)$ も 7 日間移動平均とした。この図からは、第 1 波のころは実効再生産数が比較的高かったが、第 2 波および第 3 波では実効再生産数が相対的に低いことが読み取れる。しかし、新規陽性者数に着目すると、第 1 波より第 2 波、第 3 波のほうが新規陽性者数は多くなっている。また、第 1 波と第 2 波の新規陽性者数のピークの前にそれぞれの実行再生産数のピークがあり、当然であるが、実行再生産数が新規陽性者数に対し先行性があることがわかる。

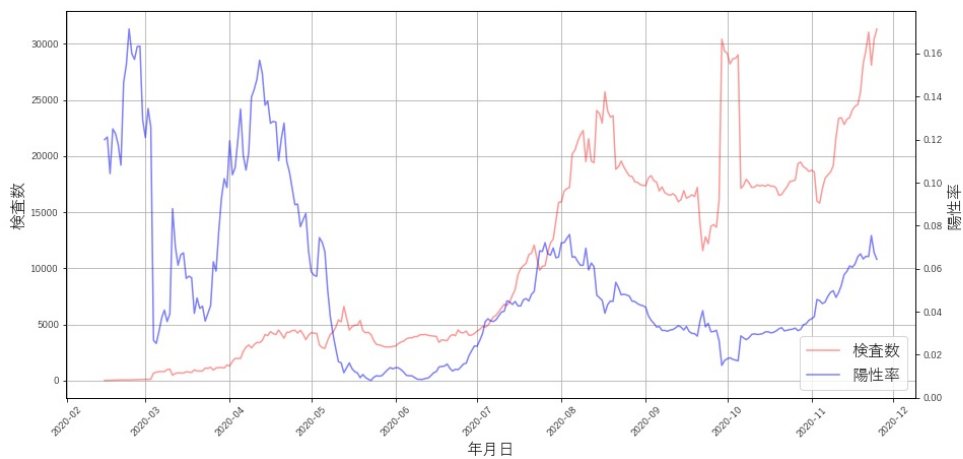


図 5: PCR 検査数および陽性率

次に、PCR 検査数である。図 5 は、PCR 検査数およびその陽性率を表している。新規陽性者数と同様に、PCR 検査数についても 7 日間移動平均をとり、陽性率については、新規陽性者数と PCR 検査数の 7 日間移動平均から算出した。まず、感染拡大当初から、日本では PCR 検査数が少ないことが指摘されてきた。グラフから、第 1 波では検査数が少なく陽性率が高いことがわかる。これは少ない検査体制でより陽性のリスクが高い人に検査を実施していたことによるものと想定される。それに対し、第 2 波以降、PCR 検査数の増加に伴い陽性率が第 1 波に比べて低いことがわかる。また、

図4と見比べると、新規陽性者数が多い時点では陽性率が高い傾向があることがわかる。また特徴的なのは、2020年10月1日付近にPCR検査数の大きなピークがある。おそらく、どこかの地域などで集中的にPCR検査を実施した結果だと想定される。それに伴い、新規陽性者数が若干増加しているが、陽性率は減少していることが読み取れる。これらから、潜在的感染者数は各時点の新規陽性者数だけでなく、PCR検査の実施状況に影響を受けるのではないかと想定した。このため、潜在的感染者の推定にはこのPCR検査に関するデータを活用すべきと考え、本研究では、PCR検査の陽性率に着目した。

状態方程式

本研究では、状態空間モデルにおける状態方程式を次のように定義した。

$$I(t) = \text{Normal}(\mu_I(t), \alpha_I I(t-1)) \quad (12)$$

$$\mu_I(t) = \gamma(t-1)\Re(t-1)I(t-1) - \Delta R(t) - \prod_{i=1}^5 \{1 - \gamma(t-i)\} I(t-5) + I(t-1) \quad (13)$$

$\mu_I(t)$ は、先に示した式(6)を観測データである新規陽性者数 $\Delta R(t)$ を用いて表した(式(13))ものである。ここで、SIRモデルにおける γ は回復率を表すが、本モデルでは陽性者を隔離者 $R(t)$ とみなしているため、 $\gamma(t)$ は感染者の検知率ととらえることができる。さらに、 $\gamma(t)$ を感染者の検知率としたため、無症状などで検知されず自然回復する人を別途考慮しなければならない。ここで、平均感染時間については、平均世代時間と等しく5日と仮定した。式(13)の最終項は、感染後(発症後)5日間検知されず自然回復した人数を表す。未観測データである潜在的感染者数 $I(t)$ は、平均を $\mu_I(t)$ 、分散を $\mu_I(t-1)$ の定数 α_I 倍した値とする正規分布に従う(式(12))ものと仮定した。また $I(0)$ については、平均 $\mu_I(0)$ 、分散を $\mu_I(0)$ の α_I 倍(式(14))とする正規分布から抽出し、ただし $\mu_I(0)$ は、初期新規陽性者数 $\Delta R(0)$ の3倍(式(15))と仮定した。

$$I(0) = \text{Normal}(\mu_I(0), \alpha_I \mu_I(0)) \quad (14)$$

$$\mu_I(0) = 3\Delta R(0) \quad (15)$$

観測方程式

観測方程式は次のとおりとした。

$$\Delta R(t) = \text{Poisson}(\lambda(t)) \quad (16)$$

$$\lambda(t) = \gamma(t-1)I(t-1) \quad (17)$$

新規陽性者(新規隔離者) $\Delta R(t)$ は、その時点の潜在的感染者 $I(t-1)$ に対して、検知率 $\gamma(t-1)$ で検知される人数を $\lambda(t)$ とするポアソン分布に従うものとする。ここで検知率 $\gamma(t)$ の高さは、PCR検査の陽性率によると考えた。陽性率が低ければ、十分な検査が実施できていると考えることができ、検知率は高いはずである。これをモデルに取り込むため、次のように定義した。

$$\gamma(t) = \text{Normal}(\mu_g(t), \sigma_g) \quad (18)$$

$$\mu_g(t) = \tanh\left(\frac{1}{\alpha_g p(t)}\right) \quad (19)$$

$p(t)$ は陽性率を表す。まず、検知率 $\gamma(t)$ は確率であるため、0 から 1 の値をとる。このため、各時点の平均値を表す $\mu_g(t)$ を \tanh 関数を使用し、モデル化 (式 (19)) した。陽性率 $p(t)$ は 0 から 1 の値をとり、その逆数が 0 に近づくほど $\mu_g(t)$ は 0 に近づき、大きくなるほど 1 に近づく。また、その傾きを調整するパラメータとして定数 α_g を設定した。検知率 $\gamma(t)$ は、平均が μ_g 、分散は定数 σ_g とする正規分布から抽出 (式 (18)) した。

表 1: 潜在的感染者推定の状態空間モデル (区画)

区画	内容	事前分布
$I(t)$	潜在的感染者	正規分布 (平均: $\mu_I(t)$ 分散: $\alpha_I I(t-1)$)
$R(t)$	隔離者 (陽性者)	ポアソン分布 (平均: $\lambda(t)$)

表 2: 潜在的感染者推定の状態空間モデル (パラメータ)

パラメータ	内容	事前分布
$\mu_I(t)$	潜在的感染者の平均値	0 以上の値
$\gamma(t)$	検知率 (0.1~0.9)	正規分布 (平均: $\mu_g(t)$ 分散: σ_g)
α_I	潜在的感染者の分散のパラメータ	一様分布 (0.1 ~ 1)
α_g	検知率の平均のパラメータ	一様分布 (40 ~ 70)
σ_g	検知率の分散	一様分布 (0.05 ~ 0.15)

ここで定義した状態空間モデルのパラメータについて、マルコフ連鎖モンテカルロ法 (以下、MCMC 法) を用いてベイズ推定を試みた。定数 α_I は 0.1 から 1 までの一様分布から抽出し、定数 α_g は 40 から 70 までの一様分布から抽出した。また、分散 σ_g は 0.01 から 0.15 までの一様分布から抽出した。本研究における MCMC 法では、繰り返し回数を 10,000 回とし、連鎖数を 4 とした。また、自己相関を緩和するため、3 回ごとに結果を抽出するものとし、推定当初の 100 回の結果は切り捨てることとした。MCMC 法が収束しているかどうかの判断は、Gelman-Rubin 統計量 (\hat{R}) が 1.1 未満となっているかどうかで判断した。

3 検証結果

3.1 感染拡大状況の把握

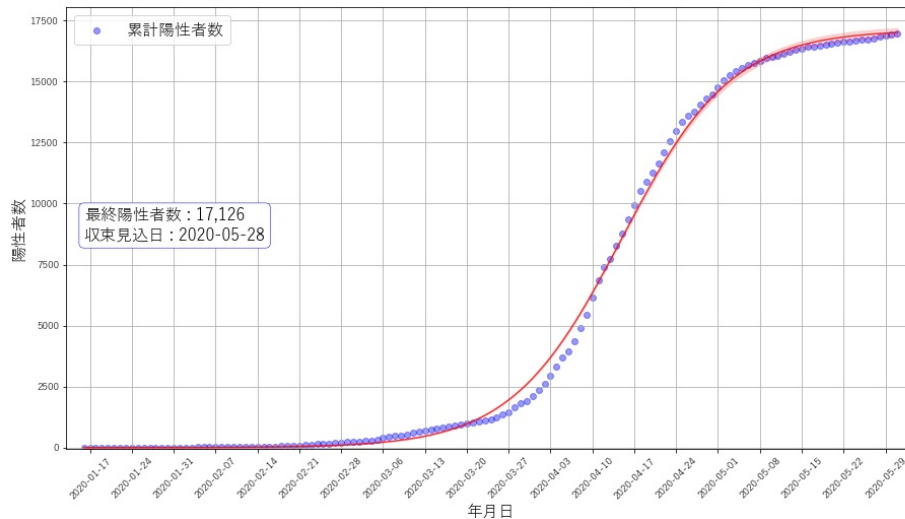


図 6: 第 1 波における感染状況のロジスティック回帰分析結果

まず、2020 年 1 月 16 日から 2020 年 5 月 31 日までの全期間の累計陽性者数に対してロジスティック回帰分析した結果を図 6 に示す。この分析から、第 1 波の最終陽性者 K は 17,126 人、2020 年 5 月 28 日に収束したと推定できる。また、感染率 r はおよそ 0.108 と推定できる。収束時点付近までのデータで推定しているため、信頼区間も幅が狭く、実データによく適合している。

しかし先に記載したとおり、感染の拡大時点までのデータによる分析では、最終陽性者数 K 、感染率 r ともに大きく異なる。各時点までのデータによる最終陽性者数 K 、感染率 r の推定結果を図 7 に示す。2020 年 2 月 16 日から 2020 年 2 月 28 日までの最終陽性者数 K がプロットされていないのは、推定値が収束しない、もしくは 10 万人以上と推定された期間である。やはり、陽性者が少ないごく初期段階では推定のブレが大きいため、2020 年 3 月以降のデータに着目する。最終陽性者数に着目すると、感染が急激に拡大した 2020 年 4 月以降、大きく上振れする。つまり、各時点までの予測は、その後の感染状況に変化なく感染が収束した場合の予測であり、早期に感染収束を予測することは困難である。逆に言えば、日々収束予測結果を観察することにより、感染状況の変化に気づくことができる。

ただし、感染率 r の解釈は難しい。新型コロナウイルスそのものの感染力は、ウイルスの変異等がない限り変化しないと想定されるが、感染状況から推定される感染率 r は感染者と未感染者の接触確率などにも影響される。また、ロジスティック回帰による手法では、最終陽性者数 K の影響も受ける。このため、時点が進むにつれて最終陽性者数の予測が増加しているのに感染率は下降しており、直感的に逆の結果となる。専門的な知見がない一般の人々でも理解しやすいよう、最終陽性者数 K に着目して議論を進める。

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、2.1 節で示した手

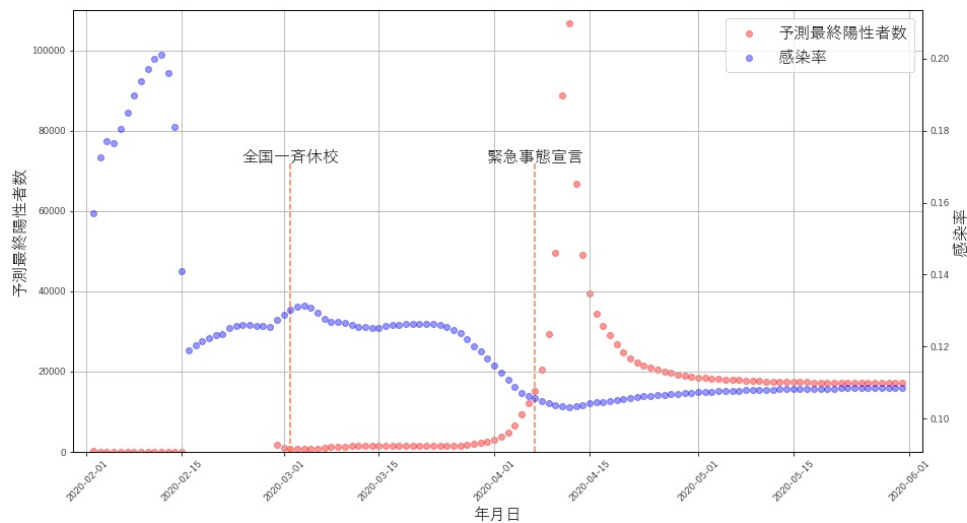


図 7: 第 1 波における感染状況の変化

法で分析を行った。その結果、変化が検出された時点のグラフを図 8~10 に示す。最初に変化が検出された 2020 年 3 月 10 日については、2020 年 3 月 2 日に小学校から高等学校までの一斉休校要請やイベント自粛要請があり、新型コロナウイルス感染症が日常生活に大きな影響を及ぼし始めたところである。この日、総理大臣官邸で第 19 回新型コロナウイルス感染症対策本部が開催され、イベント自粛の継続を呼びかけている。その効果もあってか、わずか 4 日後の 2020 年 3 月 14 日 (図 11) には最終陽性者数は 1,000 人を超えたものの、データが信頼区間内に戻った。この対応は、感染が大きく拡大する前に政府が対応を示すことにより、効果的な抑制が行えたのではないかと考えられる。

次の 2020 年 3 月 25 日は、全国的に自粛疲れとの言葉が生まれ、第 1 波の感染拡大が急速に進んだ頃である。東京都の小池都知事は「感染爆発の重大局面にある」とし、週末の外出自粛を呼びかけている。この分析結果からみても、非常に適切なタイミングでの警告であったが、結果的に対策が十分ではなかったのか、感染拡大傾向が続く。また、政府の緊急事態の宣言は 8 日後の 2020 年 4 月 7 日 (図 12) で、最終感染者の予測も約 15,000 人と 2020 年 3 月 14 日にくらべ 10 倍以上に拡大している。緊急事態宣言による経済への打撃などを意識しての結果だと考えられるが、3 月初旬にくらべ対応が遅かったことは否定できないのではないかと考えられる。その結果、データが信頼区間内に戻るのが 2020 年 4 月 15 日 (図 13) までの 21 日間を要した。

最後に 2020 年 4 月 19 日は、先ほどとは違い、データが信頼区間の下限を下回る。これは緊急事態宣言にともなう国民の自粛行動結果と考えることができる。この傾向は、感染が急拡大しようとした 4 月初めの状況もあり、5 月末まで続く。緊急事態宣言により、経済への影響は避けられなかったが、感染収束に一定の効果があつたことを示している。

このように、これら 3 時点は新型コロナウイルス感染症の第 1 波の重要な局面を抽出できている。ロジスティック回帰という非常に簡易な手法であるが、対策を検討するうえで非常に重要なインプットとなるはずである。また、これらのようなグラフを、日々公開することによる効果が大きいと考える。当時は、新規や累計陽性者のグラフはよく目にしたが、感染収束の予測についてはほとんど目にしたことがなかった。だが、2020 年 11 月 17 日より Google 社が日本における感染予測を公開して

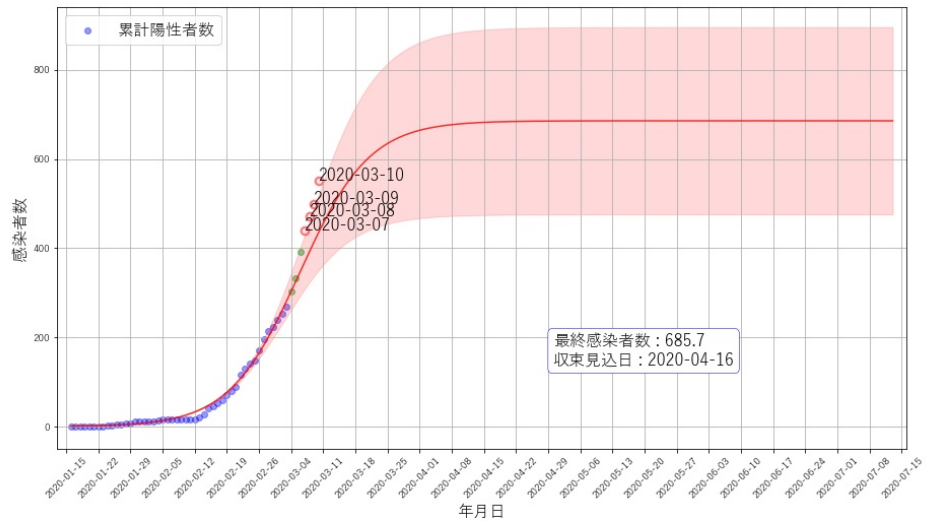


図 8: 第 1 波における感染状況 (2020 年 3 月 10 日時点)

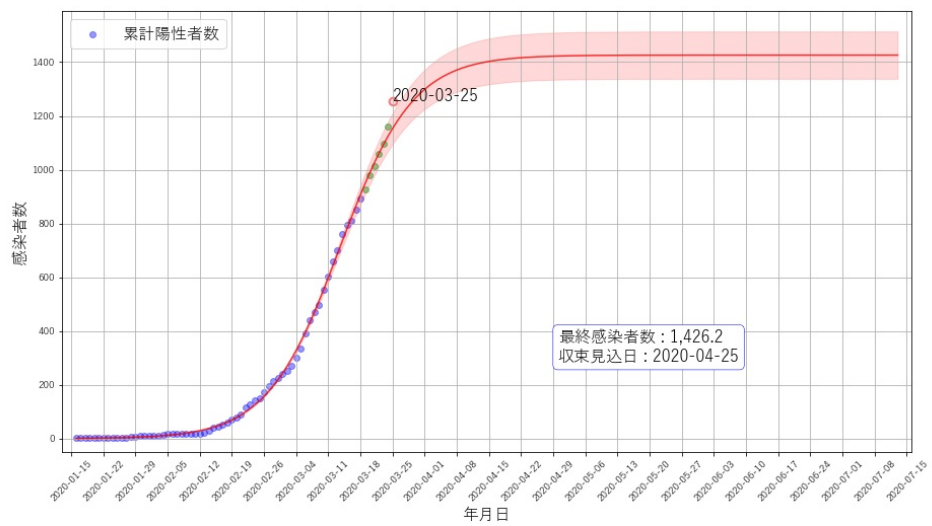


図 9: 第 1 波における感染状況 (2020 年 3 月 25 日時点)

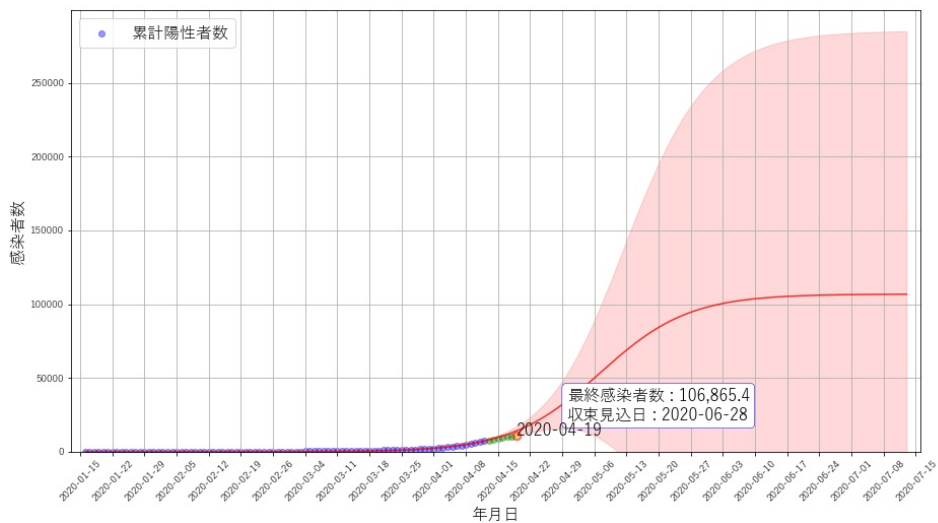


図 10: 第 1 波における感染状況 (2020 年 4 月 19 日時点)

いる [9]。これにより、現状の感染状況を客観的にとらえることができ、予測と実績を比較することにより対策の効果を検証できるようになった。こういった情報公開に本手法は活用できる。

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することもできる。そのグラフを図 14 に示す。感染状況の変化を検知した 2020 年 3 月 10 日を緑、2020 年 3 月 25 日を青、2020 年 4 月 19 日を黄で表し、その時点で予測された新規陽性者数の推移を曲線で表している。この図の曲線は、先ほどの変化点検知の手法と違い、各時点までの陽性者数のデータをすべて使用し推定していることに注意されたい。さらに、第 1 波における最大の感染拡大が予測された、2020 年 4 月 12 日を同様に赤で、期間最終日の 2020 年 5 月 31 日を黒で表している。特に、赤と黒の新規陽性者数の推移に着目すると、2020 年 4 月 12 日までの予測に基づく最大新規陽性者数約 2,755 人に対し、2020 年 5 月 31 日までの予測に基づく最大新規陽性者数約 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。このような情報を公開することにより、自身の行動がどの程度感染抑制に効果があったかを実感でき、その後の対策要請にも訴求力がますます期待される。

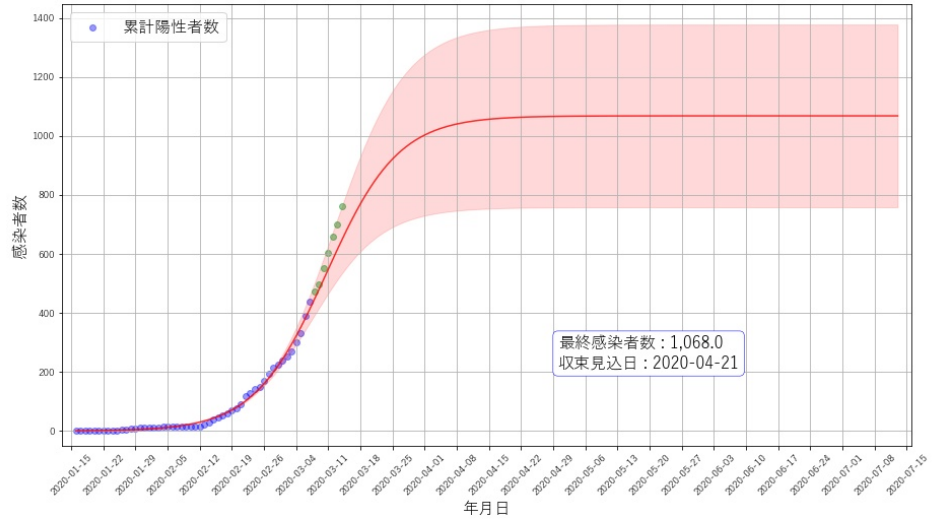


図 11: 第 1 波における感染状況 (2020 年 3 月 14 日時点)

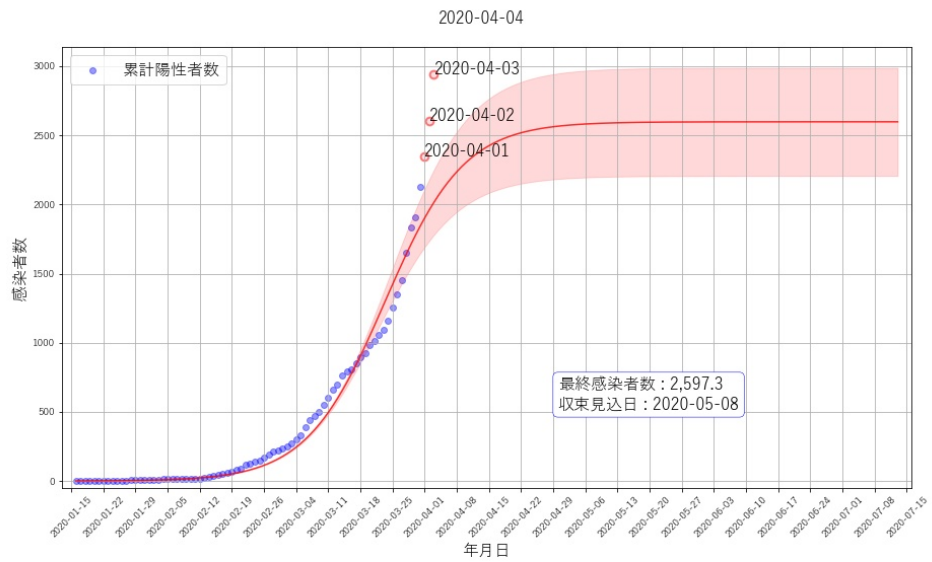


図 12: 第 1 波における感染状況 (2020 年 4 月 07 日時点)

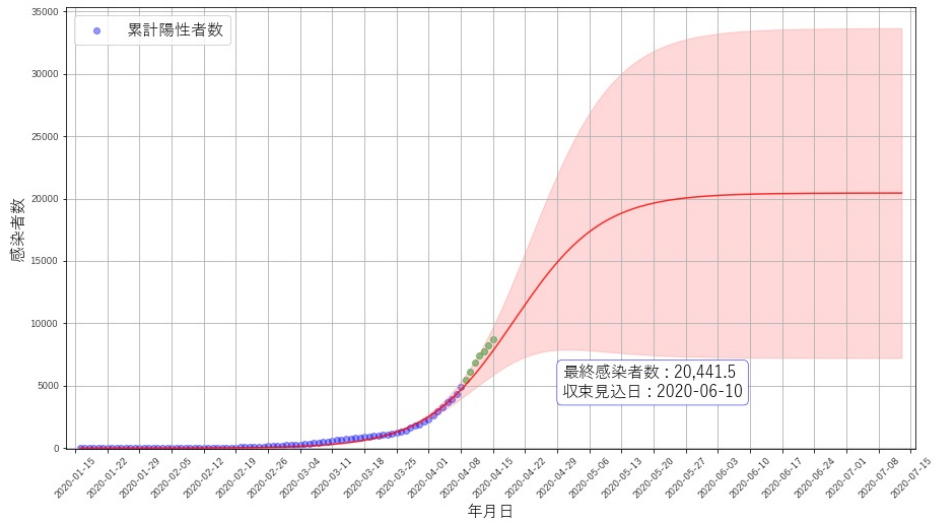


図 13: 第 1 波における感染状況 (2020 年 4 月 15 日時点)

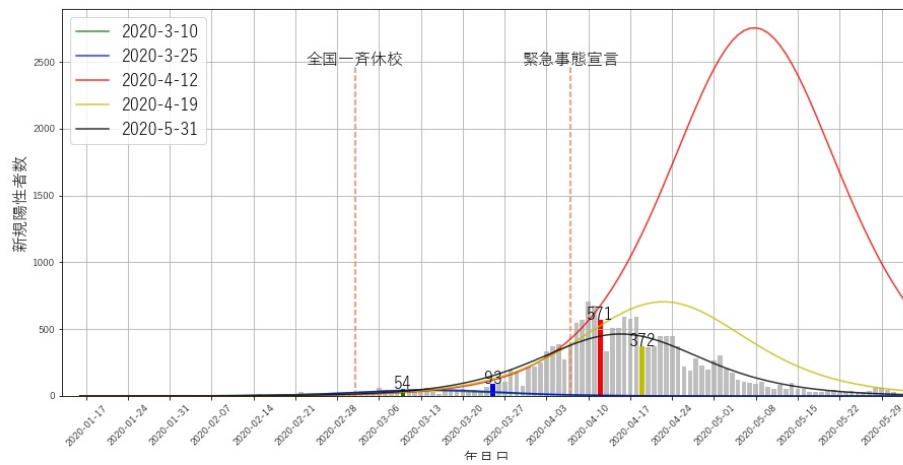


図 14: 第 1 波における緊急事態宣言の効果

3.2 潜在的感染者の推定

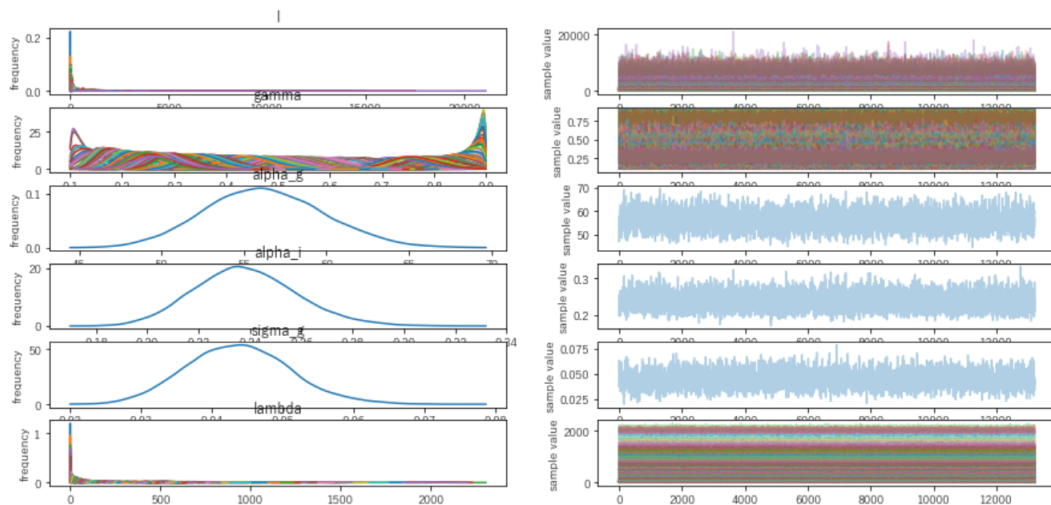


図 15: MCMC による推定結果

2.2 節で構築したモデルについて MCMC 法を実施した結果を図 15 に示す。適切に安定しており、Gelman-Rubin 統計量 (\hat{R}) はすべてのパラメータにおいて 1.1 未満で、適切に収束した。さらに、状態空間モデルによる潜在的感染者数の推定結果を図 16 に示す。日々の新規感染者数の 7 日間移動平均を黄色の棒グラフで、推定された潜在的感染者数 $I(t)$ を青の折れ線グラフで表示している。また、推定された検知率 $\gamma(t)$ を赤の折れ線グラフで、それぞれの塗りつぶしは 95% 信頼区間を表す。

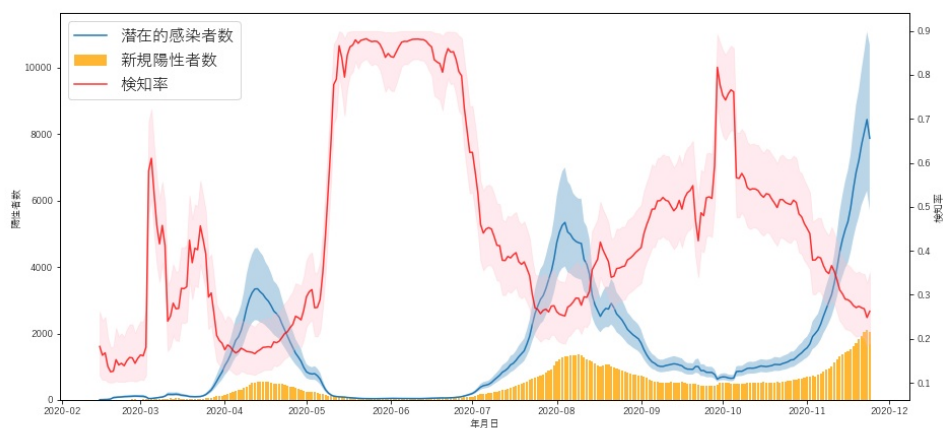


図 16: 潜在的感染者数の推定結果

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 日に約 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。同様に、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。第 1 波と第 2 波では、新規陽性者数は 2 倍程度差があるが、潜在的感染者数は 1.6 倍の差となった。第 1 波では、PCR 検査数の少なから潜在的感染者が多いことが想定されるが、それを裏付ける結果となった。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、陽性率が低い区間では、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区間では 40~50% 程度と低くなった。第 3 波の潜在的感染者は感染拡大過程であり考察対象としないが、2020 年 11 月 16 日時点ですでに第 2 波の潜在的感染者数を超過していると推測された。

ここで、累計感染者数を考える。累計推定感染者数 $I_C(t)$ は推定された各時点の潜在的感染者数 $I(t)$ と実効再生産数 $\mathfrak{R}(t)$ を用いて式 (20) のように表せる。また、各時点の累計陽性者数と累計推定感染者数を図 17 に表す。

$$I_C(t) = \sum_{i=1}^{t-1} \gamma(i)\mathfrak{R}(i)I(i) \quad (20)$$

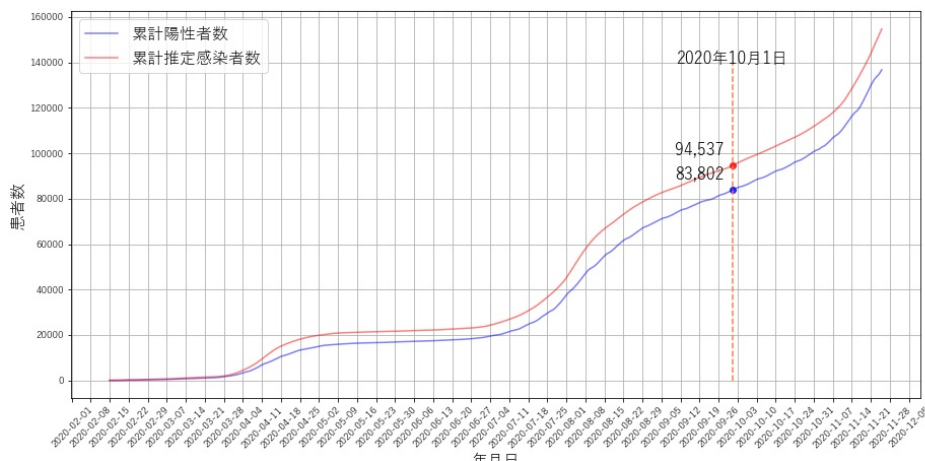


図 17: 累計推定感染者数の推定結果

森ら (2020) は兵庫県内の 5 病院、1 施設から提供を受けた 10,377 人の血清中における新型コロナウイルスの抗体の有無を調べ、中和抗体の保有率が約 0.15% であったと報告している [10]。本モデルで推定された 2020 年 10 月 1 日までの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。これは森らの研究結果より小さい結果となった。本研究においては、無症状の感染者等についての考察は行っていないことなどが影響しているかもしれないが、それでも想定以上に日本では感染者が少ないこと示している。

4 まとめと今後の課題

感染拡大状況の把握のため、感染状況の変化をいち早く察知し、それを可視化する手法について提案を行った。各時点の1週間前までの新規陽性者数のデータをもとにロジスティック回帰分析を行い、信頼区間を算出し、直近1週間の新規陽性者数のデータが信頼区間を外れた場合に変化があったものとした。その結果、感染拡大の第1波では、2020年3月10日および2020年3月25日に感染拡大の変化があったことを示すことができた。また、2020年4月19日に感染収束の変化があり、緊急事態宣言によって最大新規感染者数を約83%抑制することができたと推定された。本研究では、第1波を対象に分析を行ったが、第2波、第3波についても混合ロジスティックモデルを構築すれば同様の手法が活用できるのではないかと想定される。しかし、2020年10月1日ごろの第2波と第3波の境界では、第2波の感染が収束するまえに第3波に襲われている。こういった状況でも同様の手法で変化点が検知できるか研究を進める価値があると考ええる。

また、潜在的感染者の推定では、2020年10月1日までの累計で約94,537人の感染者が存在したことを示すことができた。2020年10月1日までの累計陽性者数の約1.13倍の感染者数となるが、日本の人口に占める割合は約0.08%と非常に低く、今後も多くの人が新型コロナウイルスに感染するリスクがあることが示された。本研究では、PCR検査の陽性率に着目したが、人の移動の増減や接触確率、陽性者に対する無症状者の割合などが考慮できていない。また、感染者が検知されずに回復するまでの日数を平均世代時間に合わせて5日と仮定した。これらをさらにモデル化することで、より感染状況の実体を表すことができるかもしれない。実際の感染症流行では様々な要素（パラメータ）が影響するはずであり、さらにモデルを改善する余地があると考ええる。

最後に、本研究においては様々な新型コロナウイルス感染症に関する分析を試みたが、想定するような結果を得ることができなかった分析も多数あった。分析を試みる中で、正確で有益な情報の重要性を非常に強く感じた。感染症のデータは、個人情報にも関連し正確な情報がなかなか入手できない。公開の範囲やその仕組み自体も手探りであった。日本においては、約100年前のスペイン風邪の流行以来の感染症であり、すべてが手探りなのは当然である。そんな中で、少しでも教訓が残せるよう、論文の執筆を試みた。本論文作成時点でも、新型コロナウイルス感染症の流行は収まっていない。今回の新型コロナウイルス感染症が早期に収束し、また100年後かもしれない新たな感染症発生時の対策に生かせるよう切に願う。

謝辞

本研究にあたり、日頃より適切な御指導をいただき、また貴重なご意見をいただきました杉本知之教授ならびに田中琢真准教授に深く感謝いたします。また、本研究過程および本稿作成全過程を通じて詳細にわたり御指導いただきました姫野哲人准教授に深く感謝いたします。

多くの御意見、御協力をいただきました姫野研究室の皆様に深く御礼申し上げます。

参考文献

- [1] 厚生労働省, ”オープンデータ”
<https://www.mhlw.go.jp/stf/covid-19/open-data.html> (参照 2020/11/26)
- [2] Nakano, T. and Ikeda, Y. (2020), Novel Indicator to Ascertain the Status and Trend of COVID-19 Spread: Modeling Study, *Journal of Medical Internet Research* **22**, e20144.
- [3] Kermack, W. O. and McKendrick A. G. (1927), A Contribution to the Mathematical Theory of Epidemics, *Proceedings of the Royal Society of London. Series A* **115**, 700-721.
- [4] 稲葉 寿 (2008), ”微分方程式と感染症数理” 数理科学 **538**, 1-7.
- [5] Thomas, V. I. (2020), Public Health Measures and the Reproduction Number of SARS-CoV-2, *Journal of the American Medical Association* **323**, 2186-2187
- [6] Cori, A., Ferguson, N.M., Fraser, C. and Cauchemez, S. (2013), A New Framework and Software to Estimate Time-Varying Reproduction Numbers During Epidemics, *American Journal of Epidemiology* **178**, 1505-1512
- [7] 西浦 博 (2020/5/22) , ”実効再生産数とその周辺”
<https://github.com/contactmodel/COVID19-Japan-Reff>
- [8] 荻原 和樹 (東洋経済オンライン編集部), ”新型コロナウイルス 国内感染の状況” 東洋経済オンライン.
<https://toyokeizai.net/sp/visual/tko/covid19/> (参照 2020/11/26)
- [9] Google 社, ”COVID-19 感染予測 (日本版)” *COVID-19 Public Forecasts*.
<https://datastudio.google.com/u/0/reporting/8224d512-a76e-4d38-91c1-935ba119eb8f/page/ncZpB> (参照 2020/12/15)
- [10] 神戸大学, ”兵庫県における新型コロナウイルス大規模血清疫学調査” *Research at kobe*.
https://www.kobe-u.ac.jp/research_at_kobe/NEWS/collaborations/2020_12_17_02.html (参照 2020/12/29)