

パンデミック SIR 方程式の確率微分方程式解法について

牧 允 皓^{†1} 廣 瀬 英 雄^{†2}

今日、感染症の拡大は人類にとって大きな問題になっている。2003 年の SARS, 2009 年の新型インフルエンザ, 2010 年の口蹄疫など、まだ記憶に新しい。感染症の流行によってもたらされる被害は、人命の危機、経済的な打撃など様々であるが、この感染拡大を事前に予測し、備えることができれば被害を抑えられる可能性がある。これまで、感染拡大予測法のモデルは多く研究されているが、ここでは、連立微分方程式からなる SIR モデルに確率項を加えた新しいモデルを提案する。これによって、予測値に信頼度を与えることができる。

A Method to Solve the Stochastic Differential Equation for SIR type Pandemic Analysis

YOSHIHIRO MAKI^{†1} and HIDEO HIROSE^{†2}

Nowadays, infectious disease spread has become one of the serious issues. We can remind SARS in 2003, new-type influenza in 2009, and foot-and-mouth disease in 2010. If we can predict the disease spread and prepare to that, we may avoid disasters or reduce damages. In this paper, we propose a method to solve the stochastic differential equation for SIR type pandemic analysis. Using this method, we can provide the reliability to predicted values.

1. はじめに

感染症などの病気は人類にとって重要な問題である。SARS, 新型インフルエンザ, 近年では口蹄疫などが社会を騒がせていた。これらの感染症によって経済的打撃を受けたり、最

悪の場合、多くの命が奪われている。感染症などの爆発的流行、いわゆるパンデミックの研究は様々な手法が報告されており、感染拡大を予測することや拡大経過を評価することは重要である。

拡大予測の手法には、常微分方程式を用いたモデル⁶⁾、不完全データを用いてモデル化された truncated モデル¹⁾、また構築された仮想空間のなかでエージェントを動かすマルチエージェントシステム (MAS)³⁾ などが挙げられる。それらの手法にはそれぞれ特徴があり、例えば SIR モデルでは、最も単純な場合、ある時刻までに得られたデータを差分方程式にあてはめることでパラメータの推定値を求め、それを使ってその後の予測を行うことができる。しかし SIR モデルでは、truncated モデルのような信頼区間についてあまり言及されていない。それは、連立微分方程式で定義された SIR モデルでは解析的に信頼区間を計算することが極めて難しいと考えられるからである。感染拡大の予測においては、収束値だけでなく、その信頼度を求めることも重要である。推定された結果がどの程度の信頼区間を持つかによって、ワクチン投与などの対策も変わってくる場合があると考えられる。

そこで、ここでは SIR モデルを確率微分方程式 (SDE: Stochastic Differential Equation) によって拡張し、予測収束値の確かさを表す信頼区間を求めることができる数値解法を提案する。確率微分方程式は株価などの経済の世界や、医療の世界で積極的に研究されており、SIR モデルの拡張が実現すれば、新しい展開が得られる。

2. パンデミック拡大の予測モデル

ここではパンデミック解析においてよく使われている二つの手法を概観する。一つは truncated モデル、もう一つは SIR モデルである。

2.1 truncated モデル

ここでは、ある時刻までに報告されている感染者数からその後の感染者数の増加を予測する truncated モデルについて述べる。パンデミック解析では、感染者が日別に報告されているデータが多いことから、truncated データ、grouped データの不完全データとして取り扱う。このとき、尤度関数は式 (1) で与えられる。ここに、打ち切り時間 (感染症では観測データの最終日) を T 、期間 T_{i-1} から T_i までの観測数を n_i 、そしてグループの数を k としている。

$$L_T(\theta) = \prod_{i=1}^k \left\{ \frac{F(T_i) - F(T_{i-1})}{F(T)} \right\}^{n_i} \quad (1)$$

^{†1} 九州工業大学大学院情報工学府
Graduate School of Science and Systems Engineering, Kyushu Institute of Technology
^{†2} 九州工業大学大学院情報工学研究院
School of Science and Systems Engineering, Kyushu Institute of Technology

パラメトリック問題としてパラメータ推定を行うにあたり、適切なデータの背後分布を仮定するが、パンデミック解析においては、growth model であるシグモイド曲線（ロジスティック関数の一つ）がよく用いられる。形状パラメータを有する一般のロジスティック関数は式（2）で与えられる。

$$F(t; K, \mu, \sigma, \beta) = K \left[1 + \exp\left(-\frac{t - \mu}{\sigma}\right) \right]^{-\beta} \quad (2)$$

ここに、変数 t は時刻を表し、パラメータは μ, σ, β の 3 つである。 K は収束値である。パラメータの推定には、最尤法を用いる。truncated モデルでは、感染拡大の終盤にならないと安定して正確な推定を求めることができないことが報告されている。

2.2 SIR モデル

SIR モデルは、マルサスの人口論に用いられる微分方程式を拡張したケルマック・マッケンドリックモデルによるもので、式（3）の 3 つの連立微分方程式で表される。

$$dS(t)/dt = -\beta S(t)I(t) \quad (3)$$

$$dI(t)/dt = \beta S(t)I(t) - \gamma I(t)$$

$$dR(t)/dt = \gamma I(t)$$

ここに、 $S(t), I(t), R(t)$ は時刻 t における感受性数 (susceptible), 感染数 (infected), 除外数 (removed) を意味する。 S はこれから感染の可能性がある集合で、 I は感染しており S に感染させる可能性がある集団で、除外数 R は回復また死亡により感染する、もしくは感染させる可能性のない集団を意味する。つまり、このモデルでは一度感染し除外されると、死亡以外では免疫をもち再び感染しないことを前提としている。式（3）のパラメータ β, γ はそれぞれ感染率、除外率である。式（3）から、任意の時刻において $N = S + I + R$ (定数) となる。ここで N は人口総数を意味する。

これらのパラメータを推定するには前進差分、後退差分などが使われることが多く、ここでは後退差分を採用した。後退差分は式（4）で与えられ、時刻 t における β, γ が得られる。

$$\beta(t) = \frac{S(t-1) - S(t)}{S(t)I(t)}, \quad \gamma(t) = \frac{R(t) - R(t-1)}{I(t)} \quad (4)$$

観測データには、感染数しか報告されないことが多いため、感染から除外に移る移行日には病理観測結果を使うのが正しいが、ここでは簡便のため適当にいくつか決めて使っている。その方法は、1) まず観測データを移行日の日数分ずらして R に格納する。2) I には移行日をずらさずに一日あたりの感染者を格納していく。3) 最後に S は総数 N と I, R と

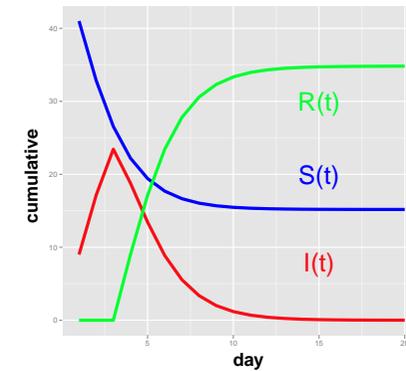


図 1 SIR モデルの一例
Fig.1 A Simulation Case of SIR Model

の差を格納していく。この行程によって、観測データから SIR のデータを作成することができる。これら値の平均を推定値とする。

β, γ のパラメータが推定されると、観測値が得られた最終日を初期値とした感染拡大のシミュレーションが可能になる。シミュレーション手法にはルンゲクッタ法を用いる。一例として、 $\beta = 0.02, \gamma = 0.9, N = 500$ とし、初期値である一日目の感染者を 0.001 としている（図 1）。

3. 確率微分方程式を用いた解法

確率微分方程式とは、確率過程が含まれる微分方程式を意味し、後述する SIR へ適用する都合でここでは微分形で書く。

$$dX_t = b(t, X_t)dt + \sigma(t, X_t)dW_t \quad (5)$$

b をドリフト係数、 σ を拡散係数といい、 W_t はホワイトノイズである。

3.1 SIR モデルへの適用

SIR モデルの 3 式の和が一定であることから、そのうちの 2 式だけについて独立な確率項を入れればよい。このことから、SIR モデルの確率微分方程式による拡張はいずれかの 2 式に拡散係数付与できれば誤差を確率的モデルとして表現できる。ここではまず、SIR モデルの $I(t)$ に確率項を取り入れ、SIR モデルの信頼区間に対する数値解法の基礎モデルを提案する。

$$dI_t = \{\beta S(t) + \gamma\}I_t dt + \alpha I_t dW_t \quad (6)$$

ここで W_t は

$$W_{t_i} = \frac{w(t_{i+1}) - w(t_i)}{t_{i+1} - t_i}$$

となる 1 次元ウィーナー過程の増加量を表し, $w(t_{i+1}) - w(t_i)$ は 1 次元の定義に従い, 平均 0, 分散 $t_{i+1} - t_i$ の正規分布に従う. $t_{i+1} - t_i$ は時間増加量である. 疑似正規乱数 ξ_i を持ちれば式 (7) を得る.

$$w(t_{i+1}) - w(t_i) = \sqrt{t_{i+1} - t_i} \xi_i \quad (7)$$

ここで伊藤の公式より, 式 (6) の両辺を 2 乗したときに次の関係式が導かれる.

$$(\alpha \xi_i)^2 = \frac{\{I(t_{i+1}) - I(t_i)\}^2}{(t_{i+1} - t_i) I^2(t_i)}$$

右辺の平方根は平均 0, 分散 α^2 の正規分布に従うので最尤推定法より推定値 $\hat{\alpha}^2$ は式 (8) のようにドリフト係数に依存せず求まる. ここで n とは観測データの数である.

$$\hat{\alpha}^2 = \frac{1}{n-1} \sum_{i=1}^{n-1} \frac{\{I(t_{i+1}) - I(t_i)\}^2}{(t_{i+1} - t_i) I^2(t_i)} \quad (8)$$

次にドリフト係数のパラメータ推定方法について述べる. 確率微分方程式ではドリフト係数のパラメータ推定に最尤推定法を用いることがあるが, ここでは SIR モデルが連立微分方程式のまま, 尤度関数を最大化するパラメータ推定を行うことは非常に困難である. そこで, 従来の後退差分を用いる. この推定方法は, 式 (3) の $dS(t)/dt$ と $dR(t)/dt$ の式を変形して得られるので, 今回のように I の式のみ拡散係数を取り入れた場合は従来と同じ推定法が使える.

3.2 SIR モデルでの数値解法

信頼区間を求める方法については, いくつか解析的な方法が報告されている. 例えば, ロジスティック関数をドリフト係数としたモデルに信頼区間を示した報告がある. しかし SIR モデルのような連立微分方程式では, 同様な方法で解析的に解くことは非常に困難である.

そこで, ここでは疑似乱数をつかい, リサンプリングによって信頼区間を示す手法を提案する. 確率微分方程式の拡散係数を疑似乱数によって計算すると, シミュレーションは従来のもとは違い, 毎回違う軌道を通る. そのシミュレーションを実現するため, ここでは確率微分方程式にもちいられる数値解法を述べる. 既に数値解法については多くの研究がなさ

表 1 シミュレーション設定
Table 1 Condition for Simulation

β	γ	α	N	打切り日 T	試行回数	初期値
$3.0 \cdot 10^{-5}$	0.3	0.3	10000	30	100	10

れており, ここではそれらの精度や特性については比較しない. 手法を固定し, SIR モデルでの有用性を検証する.

今回, 数値解法として用いるのは, 修正ホイン法である. これは安定性の面で確率ホイン法よりも優れているという報告がなされている⁵⁾. 修正ホイン法を式 (5) にあてはめると,

$$X_{n+1} = X_n + \frac{1}{2}[F_1 + F_2]h + \frac{1}{2}[G_1 + G_2]\xi_n\sqrt{h}$$

と書きなおすことができる. ただし, h は時間増加量を指し, 修正ホイン法において F, G は

$$\begin{aligned} F_1 &= b(t, X_t), & F_2 &= b(t, X_n + F_1 h + G_1 \xi_n \sqrt{h}) \\ G_1 &= \sigma(t, X_t), & G_2 &= \sigma(t, X_n + F_1 h + G_1 \xi_n \sqrt{h}), \end{aligned} \quad (9)$$

もしくは

$$\begin{aligned} F_1 &= b(t, X_t), & F_2 &= b(t, X_n + F_1 h + G_2 \xi_n \sqrt{h}) \\ G_1 &= \sigma(t, X_t), & G_2 &= \sigma(t, X_n + F_1 h + G_1 \xi_n \sqrt{h}). \end{aligned} \quad (10)$$

と定義されている. 双方強い収束次数がホイン法と同じであり, 数値解法の推定結果に大きな違いが現れなかったため, ここでは式 (9) を採用している.

4. 計 算

ここでは, まず提案法によりパラメータ推定が可能であるかを調べるためにシミュレーションを行なう. 次に, 実際に観測された SARS のデータを用いて, 感染症拡大の予測ができていないかを確認する.

4.1 シミュレーション

データ生成の条件を表 1 に示す. 感染者数の初期値を 10 とし, 疑似乱数を用いてシミュレーションを 100 回行った. シミュレーションデータからパラメータを推定する際に用いたデータは $T = 30$ (日) までのデータである. シミュレーションは初期値から修正ホイン法を用いて I, R を解いていく. 各時刻の S は総数 N を仮定することで得られる.

表 2 α の推定結果
 Table 2 Results of Estimate for α

移行日	1	2	3	4	5	6	7
$E[\hat{\alpha}]$	0.31	0.29	0.28	0.28	0.27	0.27	0.27
$V[\hat{\alpha}]$	3.2e-3	2.0e-3	2.3e-3	2.5e-3	2.6e-3	2.7e-3	2.8e-3
$E[\hat{\beta}]$	1.0e-4	5.0e-5	3.4e-5	2.7e-5	2.2e-5	1.9e-5	1.7e-5
$V[\hat{\beta}]$	6.1e-13	3.1e-12	4.6e-12	5.5e-12	6.0e-12	6.4e-12	6.6e-12
$E[\hat{\gamma}]$	1.0	0.50	0.33	0.24	0.18	0.15	0.12
$V[\hat{\gamma}]$	3.7e-3	2.0e-3	1.6e-3	1.3e-3	1.1e-3	9.7e-4	8.7e-4

これら 100 ケースのシミュレーションデータを対象に、3 節までの方法を用いてパラメータ推定を行った。移行日を 1, 3, 5, 7 と変化させた推定結果を図 2-5 に示す。100 ケースのデータから推定された α のヒストグラムである。それぞれの推定値の平均、分散を表 2 にまとめている。これらの結果より、拡散係数の推定には移行日が小さい程、大きい値を推定してしまうことがわかる。また β, γ の推定値が移行日に依って変化するのは SIR モデルの性質によるもので、ここでは詳述しない。この結果より移行日を 3 日として推定した値が、真値に近いことが確認できる。

パラメータが推定されると将来予測のためのシミュレーションが可能になる。このシミュレーションを多く（後述の例では 1000 回）繰り返すことによって各時刻における推定値がシミュレーションの回数だけ得られる。これらの推定値の分布から 95% 信頼区間を求めることができる。図 6 では 10 回シミュレーションを繰り返した結果を抜粋している。95% 信頼区間を求めるときには 1000 回のシミュレーションを行なう。

4.2 SARS の場合

2003 年 3 月 17 日から同年 6 月 12 日の間に香港で流行した SARS の観測データを用いて、前述の方法により終息値の 95% 信頼区間を計算する。打ち切り日までのデータのみが観測されているとして、30 日、50 日、70 日までのデータを用いてパラメータ推定を行い、その後感染拡大予測を行なう。それを図 7-9 に示す。図の破線の曲線が従来の SIR モデルによって解かれたもので、その上下に信頼区間 95% を示す。

これらの打ち切り日と収束値の関係を 1 つのグラフにしたものが図 10 である。横軸に打ち切り日、縦軸にその日までのデータで推定された収束値を表している。感染拡大が終盤になるにつれて終息していくのがわかる。ここで収束値とは 200 日以上シミュレーションし、曲線の増減が十分に小さくなった状態での値を意味する。

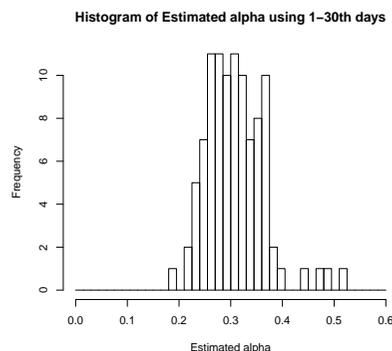


図 2 パラメータ α の推定値のヒストグラム
 (移行日: 1 日)

Fig. 2 Histogram for Estimates of Parameter α
 (step: 1 day)

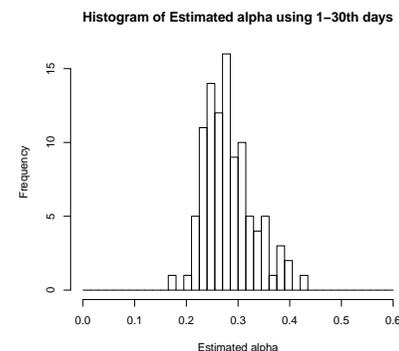


図 3 パラメータ α の推定値のヒストグラム
 (移行日: 3 日)

Fig. 3 Histogram for Estimates of Parameter α
 (step: 3 days)

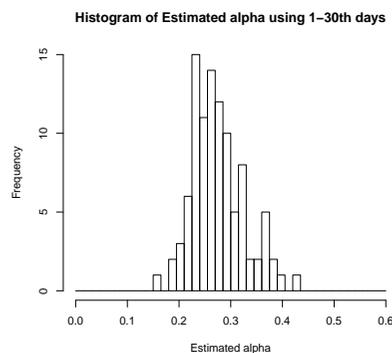


図 4 パラメータ α の推定値のヒストグラム
 (移行日: 5 日)

Fig. 4 Histogram for Estimates of Parameter α
 (step: 5 days)

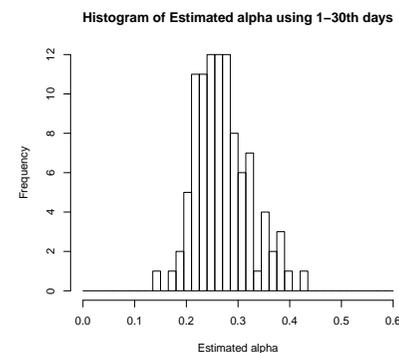


図 5 パラメータ α の推定値のヒストグラム
 (移行日: 7 日)

Fig. 5 Histogram for Estimates of Parameter α
 (step: 7 days)

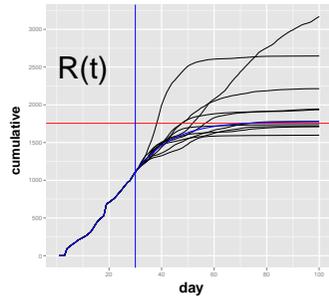


図 6 シミュレーション例
Fig. 6 An Example for SIR Simulation

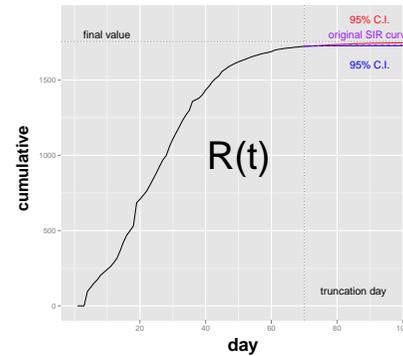


図 9 95%CI for R, 打ち切り日: 70
Fig. 9 95%CI for R, Truncation Day: 70

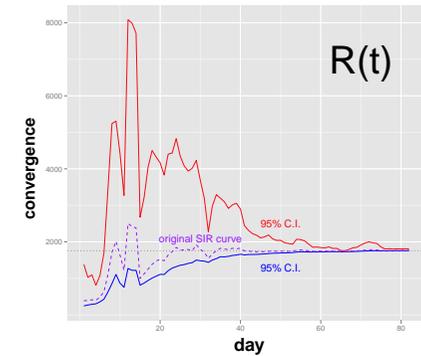


図 10 95%CI for R, 打ち切り日: 90
Fig. 10 95%CI for R, Truncation Day: 90

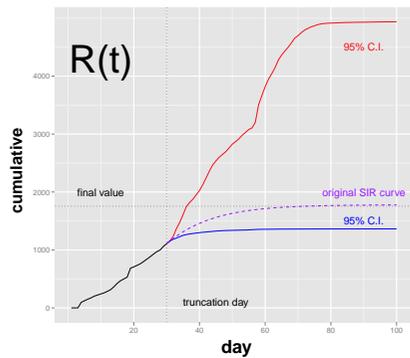


図 7 95%CI for R, 打ち切り日: 30
Fig. 7 95%CI for R, Truncation Day: 30

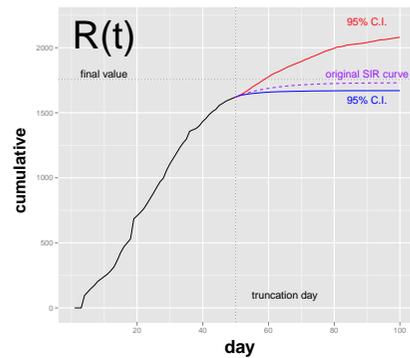


図 8 95%CI for R, 打ち切り日: 50
Fig. 8 95%CI for R, Truncation Day: 50

5. ま と め

パンデミック解析には SIR モデルに依るものが多いが、決定論的な微分方程式を用いているため予測値の信頼度は得られない。ここでは、確率微分方程式を用いて、感染者数に確率項を加えることで予測値の信頼度を与える方法を提案した。はじめに、シミュレーションを用いて提案手法の妥当性を示した。次に、実際に起こったパンデミック SARS について、確率微分方程式を適用して予測値の信頼区間を得る方法を示した。

参 考 文 献

- 1) Hirose, H.: *Estimation for the Size of Fragile Population in the Truncated and Truncated Models with Application to the Confidence Interval for the Case Fatality Ratio of SARS*, Information, Vol.12, No.1, pp.33-50 (2009).
- 2) Skiadas, C.H.: *Exact Solution of Differential Equation: Gompertz and Generalized Logistic*, Methodology and Computing in Applied Probability, Vol.12 (2009), pp.261-270.
- 3) Toyosaka, Y. and Hirose, H.: *The Consistency between the Two Kinds of Pandemic Simulations of the SEIR Model and the MAS Model*, IEICE Transactions on Fundamentals, Vol.E92-A, No.7, pp.1558-1562 (2009).
- 4) Øksendal, B.: *Stochastic Differential Equations*, Springer-Verlag Berlin Heidelberg (1998). (谷口説男 訳: 確率微分方程式, シュプリンガー・ジャパン株式会社 (2009)).
- 5) 齊藤善弘: 確率微分方程式に対する修正ホイン法, 日本応用数理学会論文誌, Vol.21, No.4, pp.323-333 (2011).
- 6) 廣瀬英雄, 松隈和広, 作村建紀: 感染症拡大予測も出るとその考察, 情報処理学会論文誌, 数理モデル化とその応用 Vol.4, No.3, pp.102-109 (2011).
- 7) WHO(World Health Organization), Cumulative Number of Reported Probable Cases of Severe Acute Respiratory Syndrome (SARS), available from (<http://www.who.int/csr/sars/country/en/index.html>) (accessed 2012-2-12)