

稲葉寿編『感染症の数理モデル』（培風館、2008）

第2章「感染症数理モデルのデータサイエンス」（西浦博 分担）の付録

（2008年7月31日改訂）

ユトレヒト大学 西浦博

はじめに

本稿は、上述の通り『感染症の数理モデル』における第2章の付録である。大学教養課程レベルの数学知識がある初学者（読者）に対しては、できるだけ余計な解析的議論を避けてデータの見方をお伝えしたかったのと、ページ数制限の都合で縮小せざるを得なかったため、私の個人的な HTML サイト上で付録（Appendix）をシェアすることにさせていただいた。本文とあわせてご利用いただきたい。この付録の A と C は編者である稲葉寿氏とのディスカッションを通じて完成したものである。同書の内容を磨きあげるために時間を惜しまず議論いただいた稲葉氏に深謝し、ここに付録をお届けしたいと思います。

付録 A. スペイン風邪（インフルエンザ）の潜伏期間推定

第2章第2節で議論した McKendrick のインフルエンザ潜伏期間の想定について議論しておこう。定量的に感染者の発症情報を分析する場合、カレンダー時刻 t だけでなく感染年齢 τ を常に明示的に考慮することは不可欠である。ここで船上の感染人口密度を $i(t, \tau)$ とすると、それは

$$i(t, \tau) = i_0(\tau - t) \frac{\Gamma(\tau)}{\Gamma(\tau - t)} \quad (1)$$

で与えられる ($\tau > 0$)。ここで $i_0(\tau)$ は乗船時初期の感染年齢分布であり、 $\Gamma(\tau)$ は感染年齢 τ の未発症生残率を表す。後者は瞬間的発症率 $\gamma(\tau)$ を用いて

$$\Gamma(\tau) = \exp\left(-\int_0^\tau \gamma(\sigma) d\sigma\right) \quad (2)$$

で与えられるものであり、潜伏期間の確率密度 $f(\tau)$ は

$$f(\tau) = \gamma(\tau) \Gamma(\tau) \quad (3)$$

で与えられる。船上での2次感染可能性を排除したデータでは $t > 0$ で $i(t, \tau) = 0$ である。

ここで、 $C(t)$ を観察された時刻 t での発症者数とすると、

$$C(t) = \int_t^\infty \gamma(\tau) i(t, \tau) d\tau \quad (4)$$

式(4)の右辺を(1)で置き換えると

$$C(t) = \int_t^\infty \gamma(\tau) \Gamma(\tau) \frac{i_0(\tau - t)}{\Gamma(\tau - t)} d\tau = \int_0^\infty f(\sigma + t) \frac{i_0(\sigma)}{\Gamma(\sigma)} d\sigma \quad (5)$$

いま、感染者の全てが発症すると仮定すると、総感染者数と総発症数は以下の関係にある：

$$\int_0^{\infty} C(t)dt = \int_0^{\infty} i_0(\tau)d\tau \quad (6)$$

式(5)と(6)を利用して、観察された船上発症者の分布（相対頻度） $c(t)$ を考えると

$$c(t) = \frac{C(t)}{\int_0^{\infty} C(t)dt} = \int_0^{\infty} \frac{f(\sigma+t)}{\Gamma(\sigma)} \left[\frac{i_0(\sigma)}{\int_0^{\infty} i_0(x)dx} \right] d\sigma \quad (7)$$

が与えられる。式(7)は、「 $i_0(\tau)$ に関する情報がなければ $c(t)$ を推定することは不可能である」ことを暗に示唆している。第2節で述べた「乗船者の感染齢の分布は、感染して未発症状態にいる者の生残率の分布と同じ」という想定の意味が理解できるだろう。

もし乗船前の単位時間辺りの感染発生数が定数 B であるとき（定常状態のとき）、 $i_0(\tau)=B\Gamma(\tau)$ を利用すると

$$c(t) = \frac{\int_0^{\infty} f(\sigma+t)d\sigma}{\int_0^{\infty} \Gamma(\tau)d\tau} = \frac{\Gamma(t)}{\int_0^{\infty} \Gamma(\tau)d\tau} \quad (8)$$

であり、これは本文の式(2.14)で与えられた生残率 $1 - F(t)$ の分布に等しくなる。このように、乗船時の未発症感染者の感染齢分布に何らかの想定が施されない限り、船上の発症データだけから発症までの平均待機時間を推定することは難しい。

付録 B. 単純分岐過程を利用した第 n 世代の感染個体数の分散

第4節で議論した単純分岐過程を利用した場合の第 n 世代の感染個体数の分散 V_n を導出しよう[40,47]（文献は本文参照）。第 n 世代の感染個体数の平均値 M_n を導出する際に、本文の式(2.63)の両辺を微分して

$$g'_n(s) = g'_{n-1}(g(s))g'(s) \quad (9)$$

を利用した。ここでは、さらに2階微分

$$g''_n(s) = E(\rho(\rho-1)s^{\rho-2}) \quad (10)$$

を利用する。式(10)より、 $g''(1) = E(\rho(\rho-1)) = E(\rho^2) - E(\rho)$ である。従って、

$$\text{Var}(\rho) = E(\rho^2) - (E(\rho))^2 = g''(1) + R_0 - R_0^2 \quad (11)$$

が与えられる。

また、本文式(2.63)の2階微分は（式(9)両辺をさらに微分すると）

$$g''_n(s) = g''_{n-1}(g(s))(g'(s))^2 + g'_{n-1}(g(s))g''(s) \quad (12)$$

である。第0世代の感染個体数 $C_0=1$ の場合、第 n 世代の感染個体数の分散 V_n は式(11)を利用して

$$V_n = g''_n(1) + M_n - M_n^2 \quad (13)$$

と書くことができる。さらに、式(12)の結果から、(13)は以下のように書き換えられる

$$\begin{aligned}
V_n &= g_{n-1}''(g(1))(g'(1))^2 + g_{n-1}'(g(1))g''(1) + M_n - M_n^2 \\
&= R_0^2(V_{n-1} - R_0^{n-1} + R_0^{2(n-1)}) + R_0^{n-1}(\text{Var}(\rho) - R_0 + R_0^2) + R_0^n - R_0^{2n} \\
&= R_0^2 V_{n-1} + R_0^{n-1} \text{Var}(\rho)
\end{aligned} \tag{14}$$

第0世代の分散 $V_0=0$ であることを用いれば、以下のように式(14)を解くことができる：

$$\begin{aligned}
V_n &= R_0^2(R_0^2 V_{n-2} + R_0^{n-2} \text{Var}(\rho)) + R_0^{n-1} \text{Var}(\rho) \\
&= (R_0^2)^2 V_{n-2} + R_0^{n-1} \text{Var}(\rho)(1 + R_0) \\
&= (R_0^2)^2(R_0^2 V_{n-3} + R_0^{n-3} \text{Var}(\rho)) + R_0^{n-1} \text{Var}(\rho)(1 + R_0) \\
&= (R_0^2)^3 V_{n-3} + R_0^{n-1} \text{Var}(\rho)(1 + R_0 + R_0^2) \\
&= \dots \\
&= (R_0^2)^n V_0 + R_0^{n-1} \text{Var}(\rho)(1 + R_0 + \dots + R_0^{n-1}) \\
&= R_0^{n-1} \text{Var}(\rho)(1 + R_0 + \dots + R_0^{n-1})
\end{aligned} \tag{15}$$

以上より、

$$V_n = \begin{cases} n \text{Var}(\rho) & (R_0 = 1) \\ R_0^{n-1} \text{Var}(\rho) \frac{R_0^n - 1}{R_0 - 1} & (R_0 \neq 1) \end{cases} \tag{16}$$

が与えられる。 $R_0 \geq 1$ の場合、分散 V_n は n が大きくなるに従って大きくなる。一方で、 $R_0 < 1$ の場合、十分に大きい n で V_n は 0 に近づく。

付録 C. 発症を基にした再生方程式

ここで、本文の式(2.103)における発症を基にした再生方程式の想定を考えよう。時刻 t での新規感染者数を $i(t)$ 、発症者数を $c(t)$ とする。 $S(t)$ を時刻 t における感受性宿主数とすると、

$$i(t) = S(t) \int_0^\infty \beta(\sigma) \Gamma(\sigma) c(t - \sigma) d\sigma \tag{17}$$

で新規感染者が与えられる。ここで、 $\beta(\sigma)$ は病期齢 σ における感染率、 $\Gamma(\sigma)$ は発症者の生存率である。式(17)の重要な想定の一つとして、病期齢 0 以前（発症前）は 2 次感染が起こらないような感染症を考えていることに注意されたい。

逆計算法にあるように $c(t)$ は $i(t)$ と潜伏期間の密度関数 $f(\tau)$ を利用して

$$c(t) = \int_0^\infty f(\tau) i(t - \tau) d\tau \tag{18}$$

で与えられる。式(18)で(17)の右辺を置き換えると

$$i(t) = S(t) \int_0^\infty \phi(s) i(t - s) ds \tag{19}$$

ここで式(19)の s は感染後の経過時間(感染齢)であり、右辺の

$$\phi(s) = \int_0^s \beta(a)\Gamma(a)f(s-a)da \quad (20)$$

は世代時間の頻度を反映しているものとする。

時刻 t の再生産数は式(20)から

$$R(t) = S(t) \int_0^\infty \phi(s)ds = S(t) \int_0^\infty \beta(\sigma)\Gamma(\sigma)d\sigma \int_0^\infty f(\tau)d\tau \quad (21)$$

式(21)は $R(t)=R_0S(t)/S(0)$ に簡略化することができ、これは Kermack-McKendrick 型モデルにある、「感染症流行の非線形性は感受性宿主の減少によって特徴付けられる」という最も重要な想定を反映している。

一方で、式(18)の右辺を式(17)で置き換えると

$$c(t) = \int_0^\infty \psi(t, \sigma)c(t-\sigma)d\sigma \quad (22)$$

ここで $\psi(t, \sigma)$ は病期齢 σ における発症間隔を反映しており

$$\psi(t, \sigma) = \int_0^\sigma \beta(\sigma-\tau)\Gamma(\sigma-\tau)f(\tau)S(t-\tau)d\tau \quad (23)$$

で与えられる。

式(23)の積分核は $S(t-\tau)$ を含んでおり、変数分離型でないことから、式(21)のように解くことが難しい。しかし、 $\beta(\tau)\Gamma(\tau)=k\delta(\tau)$ (k は定数、 δ はデルタ関数) のように 2 次感染の相対頻度が極端に単純化できるとき

$$\psi(t, \sigma) = kf(\sigma)S(t-\sigma) \quad (24)$$

であるから、これを式(22)に戻すと

$$\begin{aligned} c(t) &= \int_0^\infty kf(\sigma)S(t-\sigma)c(t-\sigma)d\sigma \\ &= \int_0^\infty R(t-\sigma)c(t-\sigma) \frac{f(\sigma)}{\int_0^\infty f(\tau)d\tau} d\sigma \end{aligned} \quad (25)$$

が与えられ、これは本文の式(2.103)と同じ「発症を基にした再生方程式」である。式(25)から気付くことは、式(24)の想定の際に「感染者は発症後ただちに 2 次感染を起こす」と仮定したことの影響で潜伏期間分布が発症間隔分布を与えるように暗黙の間に想定した点である。

ちなみに、上記の発症数の再生方程式(25)から得られる瞬間的再生産数は

$$\begin{aligned} R_{case}(t) &= \int_0^\infty \psi(t, \sigma)d\sigma \\ &= \int_0^\infty \beta(\sigma)\Gamma(\sigma)d\sigma \times \int_0^\infty f(\tau) \frac{R(t-\tau)}{\int_0^\infty \phi(x)dx} d\tau \\ &= \int_0^\infty R(t-\tau) \frac{f(\tau)}{\int_0^\infty f(x)dx} d\tau \end{aligned} \quad (26)$$

である。言い換えれば、 $R_{\text{case}}(t)$ と $f(t)$ の情報があれば上述の $R(t)$ を推定できることに他ならない。

おわりに

最後に1つの疑問に関して議論したい。

それは「アイスコーヒーはどこまでアイスコーヒーなのか」についてである。具体的に書くならば、

ナイスコーヒー、アイスモーヒー、マイルスモーヒー、あいつもういい、ナイスコーチ
コーヒーコーヒー、たんす納品、熱いコーヒー、ハイソサイティー、アイアンヒーロー

と枚挙に暇がない（個人的には上段は全て成功した経験がある）。実は、これは感染症の数理モデルにおける、ある議論と酷似している。それは、「どこまでが現実の感染症流行を課題にしたモデルで、どこからが単なる解析的あるいは定性的議論を目的としたものであるのか。」である。

同問題に関して Explicit な線引きをすることは不可能であるが、欧州と比較して日本ではそれがハッキリしすぎており、両者の間に大きな境（というか、大渓谷のようなもの。ハッキリ書くならば興味・視点と数理的なレベルなどに関する違い）を見てきた（そして両者の間に立つことは常にストレスであり、色々なことを感じさせられてきた）。現場で感染症数理を曲がりなりにも議論する者の解析学や理論的知識の欠如は全く否めない状況にある。シミュレーション科学をはじめとする数値解析の想定や最も重要な Validity について議論せずに政策側の議論だけが進んでしまうのは Strange でしかなく、個人的には本当に残念であり続けている。また、定性的研究でも現実性の議論をひどく欠いたものがあり、同分野をリードする欧州の者たちと全く異なる視点で解析的議論が進むのも少し残念だった。個人的には、まず応用範囲側にいる者たちが Public health ばかりを考えすぎずに基礎的な解析的 Insights を得るところからだと思ふ。そうでないと話が始まらない。

私の章は（稲葉さんのと併せて）、それを埋める1つの礎になればと、勝手なことを大袈裟に考えて書いたものである。たぶん、そんな章1つだけが助けになるなんていうことはないだろうけれども、そういう試みを世に残しておこうと努力された稲葉さんには本当に感謝したい。定量的研究をする立場上、少なくとも専門家（と自称する者）は数値から解析まで関連する全てのことを理解していなければならないが、それをどのように社会へ還元するか、は今後も私の教育課題における最大のテーマの1つであり続けるだろう。

この付録稿（特に「おわりに」）は同章執筆者の西浦による勝手な視点に基づくものであり、その文責は西浦にあります。また、本文では Typo が散見されます（P69L8の「麻疹」は「麻疹」のように）。根底の理屈や基礎すぎる部分は参考文献を紐解いていただければと思いますが、明らかな間違いやエラーがありました場合は西浦にご連絡下さい（nishiurah(at)aol.com で(at)を@に変えて下さい）。