

# 感染症の家庭内伝播の確率モデル： 人工的な実験環境

西浦 博<sup>†</sup>

(受付 2008 年 6 月 30 日；改訂 2008 年 8 月 19 日)

## 要 旨

感染症の流行について、分岐過程や出生死亡過程および空間構造を持つコンタクトプロセスなどを代表として、これまでに様々な確率過程を基礎にした疫学モデルが提案されてきた。統計学的推定を動機に疫学的な観察データを分析するとき、流行メカニズムを定量的に明らかにするためには、利用するモデルが伝播の異質性(複数の異なる接触パターン・様式などによって感染現象が起こること)を確実に捉えていることが望ましい。最も重要な異質性の 1 つとして家庭内伝播(household transmission)が挙げられる。家庭内は家族構成員の全てが濃厚な接触を経験する人工的な実験環境と考えられ、その観察記録は数多くの推定問題を解決するために有用な自然感染データとして利用されてきた。本稿は家庭内伝播に関する確率モデルについて、各モデルの応用とデータの見方を議論しながら、重要な疫学的想定と伝播機構の考え方を解説する。特に、家庭内の最終規模を利用した家庭内 2 次発症割合の推定手法と家庭間再生産数に基づく流行閾値の基礎理論について議論を展開する。これまでに複数の異なるモデルが提案されてきたが、いずれを利用する場合も対象とする感染症と流行データがモデルの想定を十分に満たすことを評価する作業が欠かせない。

キーワード：感染症，確率過程，疫学，流行，家庭内伝播，モデル。

## 1. はじめに

直接伝播する感染症の疫学的な観察データは他の疾病のそれと比較して非常に特異的である。最も重要な 2 つの特徴として、(1) 宿主から宿主へと伝播が起こるために 1 人の感染者の観察は他者の観察と独立でないこと、(2) 性感染症の接触者追跡調査など特別な事例を除いて感染現象や感染性の獲得は直接に観察できないこと、を常に念頭に置いてデータ分析を実施することが望ましい(西浦, 2008)。前者の特徴があるために、予防接種政策はワクチン接種個体だけでなくワクチン接種を実施した集団全体に対する感染防御能力(集団免疫)を与える。また、同じ理由のために、他疾患の疫学研究で頻用されるリスク比やオッズ比を直接に感染リスクの分析に適用することが不適切な場合も少なくない(Koopman and Longini, 1994)。結果として従属性現象(dependent happening)である感染過程のデータを処理するために、新規感染に関する非線形の流行動態を明示的に考慮した数理モデルが必要とされる(西浦・稲葉, 2006)。観察データを基にモデルのパラメータ推定を実施する場合、さらに人口学的確率性(demographic

---

<sup>†</sup> Theoretical Epidemiology, University of Utrecht, Yalelaan 7, 3584 CL, Utrecht, The Netherlands;  
h.nishiura@uu.nl

stochasticity; 感染個体数が有限および離散的であるために生じる確率性)に伴うデータ変動を加味した確率論的モデルを適切に使用することが求められる。疫学モデルの多くは、他の分野でも頻繁に利用される分岐過程や出生死亡過程および空間構造を持つコンタクトプロセスを代表として、様々な確率過程を基礎に提案されてきた。本稿では家庭内伝播に焦点を絞るために個々の基礎を詳述しないので、一般的な確率過程の基礎と感染症疫学における応用については他書を参考にさせていただきたい (Andersson and Britton, 2000; Bailey, 1964; 藤曲, 2003; Karlin and Taylor, 1975)。

疫学モデルを用いて観察データを分析するとき、明らかにしたい流行動態を適切に定量化するためには、利用するモデルが伝播の異質性(複数の異なる接触パターン・様式などによって感染現象が起こること)を確実に捉えていることが望ましい。様々な異質性に関する対処のうち、最も重要な 1 つとして家庭内伝播 (household transmission) を家庭外(コミュニティ)のそれと区別することが挙げられる。ここで、家庭(あるいは家屋; household)とは、「(血縁関係のあるなしに関わらず)1つの屋根の下で同じ生活環境を共有する個体群」と定義される (Becker, 1989)。少数の個体群が長い時間を共有する家屋は人工的に作り上げられた実験環境にほかならず、全ての家族構成員が濃密な接触を経験する特殊なデータを与える。直接伝播する感染症の多くはコミュニティよりも家庭内で伝播する確率のほうが高いことが多い。そのため、興味の対象とする人口集団において一世帯当たりの人員数や家屋数がどのように分布するかは感染症の流行閾値(ある感染症が対象集団において非ゼロの確率で大規模流行を起こすか否かを決定する閾値)を導く上で欠かせない情報と言えるだろう。また、これまでに、連鎖型二項分布モデル(chain binomial model)と称される Reed-Frost 型モデルや Greenwood 型モデルをはじめとして、家庭内伝播の観察記録は数多くの推定問題を解決する有用な自然感染データとして利用されてきた (Abbey, 1952; Greenwood, 1931)。現に、わが国の先駆的な理論疫学の専門書においても、本課題に関する古典的モデルの考え方が必ず議論されている (平山, 1958; 金光他, 1966; 阪本, 1985)。

しかし、家庭内伝播に関するモデルの考え方は古典的な連鎖型二項分布の提案以降に飛躍的に発展し、均質な接触を経験する個体群における感染症の流行理論と比較して非常に専門性の高い分野に変容した (Daley and Gani, 2001)。また、モデルの考え方と用途が多様化したため、その種類と想定は多岐にわたる。よって、初学者が家庭内伝播を適切に理解するためには、この課題に関する網羅的なチャートを与える総説が必要とされるだろう。以上のことから、本稿は家庭内伝播に関する確率モデルの基礎的な考え方を解説する。次節以降、モデルの応用と観察データの見方を交えつつ解説するが、本稿は応用事例そのものよりも各モデルの重要な疫学的想定と家庭内の伝播機構に対する考え方を習得するための基礎資料にすることを第一の目的として、様々なモデル構造に関する議論を展開する。

## 2. 連鎖型二項分布モデルと家庭内伝播

最初に、家庭内伝播に関するアイデアの基礎を得るために、19 世紀後半から 20 世紀中盤までに盛んに利用された連鎖型二項分布モデルを検討しよう。特に、同モデルを利用することによって家庭内感染データの考え方を議論する。連鎖型二項分布モデルは、元々は家庭内の感染者数でなく、特定規模の集団における時系列の感染者数(流行曲線)を描写するために提案された。そこで、まずは集団の流行動態を説明するモデルに関して最も原始的な構造を紹介する。

### 2.1 En'ko による麻疹の流行モデル

最も古い疫学モデルは Daniel Bernoulli や Theophil Lotz などによる天然痘流行の描写を代表として 18-19 世紀に提案されたが、それらは年齢別既感染者割合や幾何級数的な感染者数の

増殖を説明する線形のモデルであった (Dietz and Heesterbeek, 2002; Nishiura et al., 2006). 感受性宿主の減少によって特徴付けられる非線形の新規感染に関する動態は 19 世紀後半まで明らかにされず, それは 1889 年にロシア人 Piotr Dmitrievich En'ko [1844-1916] が最も単純な連鎖型二項分布を提案したことに始まる. 原文はロシア語で執筆されているが, En'ko の没後 100 年を機会に同文は Klaus Dietz によって英訳出版されている (En'ko, 1989). 1870 年代のサントペテルスブルグの麻疹 (はしか) の流行を説明するために, En'ko は以下の反復方程式を用いた.

$$\begin{aligned}
 (2.1) \quad & I_{t+1} = S_t \left\{ 1 - \left( 1 - \frac{I_t}{N_t - 1} \right)^{kN_t} \right\} \\
 & S_{t+1} = S_t \left( 1 - \frac{I_t}{N_t - 1} \right)^{kN_t} \\
 & N_{t+1} = N_t - I_t
 \end{aligned}$$

ここで  $I_t$  は単位時間  $t$  において観察される感染者数,  $S_t$  は同期間の感受性宿主の数,  $N_t$  は総人口である. パラメータ  $k$  は 1 人の感受性宿主が単位時間あたりに経験する接触回数であり,  $A_t = kN_t$  は対象人口における総接触回数を与える. 式 (2.1) の流行モデルにおける最も重要な想定を要約すれば, En'ko は感染者が 1 つの単位時間の後に直ちに隔離されるために人口から除外されると想定し, さらに, 接触回数は総人口に比例すると考えた (Dietz, 1988).

流行開始前の総人口  $N_0$  が既知であり (例えば 400 人としよう), 初期感染者数  $I_0 (=1$  人とする) が観察されたとすると, 初期の感受性宿主数  $S_0$  とパラメータ  $k$  の 2 つによって流行動態が説明できる. そして,  $I_t$  は単位時間  $t$  に観察された感染者数の期待値を与える. 図 1 は 1874 年の麻疹の流行曲線に対して式 (2.1) を適用した結果である.  $S_0$  および  $k$  はそれぞれ 133 人および 0.9 と推定された. 原文では, 同モデルを利用して他年度の流行曲線も分析しているので参考にしていただきたい (En'ko, 1989). 観察値と予測値に大きな乖離が見られる年度もあったが, En'ko はそれを個体間の感受性に関する異質性に帰することができると議論した.

En'ko のモデル構造は約 40 年後に提案される Reed-Frost 型モデルを既に予期していた. ここで, 総人口数に比例する総接触回数  $A_t$  を利用せずに, 接触回数の確率母関数  $f$  を使って反復方程式 (2.1) を考えてみよう:

$$(2.2) \quad f(x) = \sum_{i=1}^{\infty} p_i x^i.$$

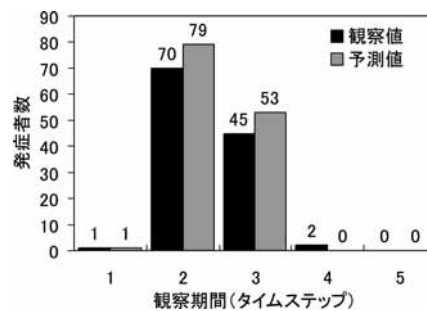


図 1. P. D. En'ko による 1874 年サントペテルスブルグの麻疹流行モデル. 横軸は単位期間, 縦軸は各期間で観察された発症者数を表す. 人口  $N_0 = 400$  を想定し, 式 (2.1) を流行曲線に適用することによって  $S_0 = 133, k = 0.9$  と推定された (Dietz, 1988).

ここで  $p_i$  は 1 人の感受性宿主が単位時間あたりに  $i$  回の接触を経験する確率である. En'ko の反復方程式 (2.1) における感染過程は, 1 人の感受性宿主が感染性宿主と接触を経験する回数に依存する. すなわち,  $I_{t+1}$  は以下のように書き換えて一般化することができる.

$$(2.3) \quad I_{t+1} = S_t \left\{ 1 - f \left( 1 - \frac{I_t}{N_t - 1} \right) \right\}$$

En'ko のモデルと同様に, Reed-Frost 型モデルも単位時間当たりの感染者数 (流行曲線) を描写することに用いられるが, その基本構造は式 (2.3) の特別な場合である. つまり, En'ko が想定した隔離を無視して (総人口数は時刻に独立と想定し), さらに接触回数が平均値を  $K$  とするポアソン分布に従うとすると

$$(2.4) \quad I_{t+1} = S_t \left\{ 1 - \exp \left( -\frac{KI_t}{N-1} \right) \right\}$$

を得る. ここで

$$(2.5) \quad p = 1 - \exp \left( -\frac{K}{N-1} \right)$$

とすると, 多くの疫学教科書に説明されている Reed-Frost 型モデル,

$$(2.6) \quad I_{t+1} = S_t \{ 1 - (1-p)^{I_t} \}$$

を得る (Abbey, 1952). 式 (2.5) より,  $p$  は 1 人の感受性宿主が単位時間あたりに他の 1 個体と少なくとも 1 回の接触を経験する確率である. より明確に Reed-Frost 型モデルを書いておこう. 式 (2.6) で明らかなように, 単位時間  $t$  毎に感染者数  $I_t$  が異なるため, 一般的な Reed-Frost 型モデルは単位時間当たりの感染者数に条件付けした 1 人の感受性宿主の接触確率  $p_{I_t}$  を用いる. 単位時間  $t$  における感受性宿主および感染性宿主の数を  $S_t$  および  $I_t$  とする. これらの観察値が与えられたとき, 単位時間  $t+1$  における感染者数  $I_{t+1}$  の条件付き分布は以下の二項分布に従う:

$$(2.7) \quad \Pr(I_{t+1} = x | S_t = s, I_t = i) = \frac{s!}{x!(s-x)!} p_i^x q_i^{s-x}.$$

ここで  $q_a = 1 - p_a$  であり,  $a$  名の感染者が存在する集団において感受性宿主が単位時間あたりに接触を経験しない確率を意味する. 式 (2.6) から明らかなように Reed-Frost 型モデルは  $q_a = q^a$  ( $q$  は時刻に独立な接触を回避する確率) を想定している. 流行初期における感染者数と感受性宿主数および世代に独立な 1 人の感受性宿主当たりの接触確率  $p$  が与えられたとき, 式 (2.7) は全ての単位時間  $t$  において  $t$  より前の観察データに条件付けした感染者数の分布が得られることを意味する. つまり, 時間  $t$  における感染者数観察の尤度は流行開始時 0 から  $t-1$  までについて, 式 (2.7) と同様の二項分布の積をとることで与えられる:

$$(2.8) \quad \Pr(I_0 \rightarrow I_1 \rightarrow \dots \rightarrow I_t) = \frac{S_0!}{I_0! \dots I_t! S_t!} \prod_{k=0}^{t-1} p_{I_k}^{I_{k+1}} q_{I_k}^{S_{k+1}}.$$

式 (2.8) に見られる各単位時間の観察に対する条件付き尤度の特徴から, Reed-Frost 型モデルや類似のモデルは (上述の En'ko のモデルも含めて) 連鎖型二項分布モデル (chain binomial model) と称される. Reed と Frost は提唱者の氏名 (Lowell Reed [1886-1966] および Wade Hampton Frost [1880-1938]) であるが, 提唱者は 1928 年に Johns Hopkins 大学において同モデルに関する講義を行なったにも関わらず, それを論文として出版・報告しなかったことが知られている (関連する Frost の研究記録に関しては Maxcy, 1941 や Daniel, 2004 が詳しい). 初回講義の 20

年以上の後に、このモデルは同大学の Helen Abbey (1952) によって発表された。本稿は家庭内伝播に着目するために、これ以上は Reed-Frost 型モデルの流行曲線を対象とする応用について説明しないが、同課題は Fine (1977) によって詳説されているので参考にさせていただきたい。また、古典的な理論ではあるものの、上述のように非線形の流行動態を離散的な反復方程式によって描写できるため、同様のモデルは簡便かつ定量的な発展性に富むことで知られている。最近でも、背景にある想定 of 妥当性に注意しつつ連鎖型二項分布モデルを応用した研究が報告されている (Longini et al., 2004; Tsutsui et al., 2003)。

## 2.2 家庭内伝播への応用

Reed-Frost 型モデルは、家庭内伝播の観察データを説明するモデルとしても利用することができる。特に、このモデルは家庭内において「1 人の感染者が侵入した場合」に、「最終的に何名の者が感染を経験するか」を表す最終規模 (final size) の分布を与える。家庭内伝播に関する疫学データの 1 つとして、家庭内の最終規模 (特定の世帯人員数の家屋を対象に、流行終息までに何名が感染したのかを表す指標) が数多く観察されてきた。最終規模の考え方を理解するために、まずは具体的な数値例を考えてみよう。

例えば、流行開始前に 5 人の感受性を有する者がいる家庭において、1 人がコミュニティで感染し ( $I_0 = 1$ )、残りの 4 人に家庭内で伝播が起こるかどうかが ( $S_0 = 4$ ) を分析したいものとする。この家庭内における最終規模は「初期感染者 1 人の感染のみ (2 次感染者は 0 人)」、「1 人の 2 次感染者が観察される」あるいは「2 人の 2 次感染者が観察される」(同様に 2 次感染者は最大 4 人まで可能) という 5 通りの組合せがある。この 5 人家庭における最終規模の分布を分析するためには、各最終規模を観察する確率を与える家庭内の流行連鎖 (epidemic chain) を個々に検討すれば良い。その際、式 (2.8) における単位時間は家庭内の感染世代で置き換える (感染世代とは家庭内の感染順序を表す; 初期感染者は第 0 世代で初期感染者によって生み出された 2 次感染者は第 1 世代である)。例えば、最終的に 3 人の 2 次感染者が観察される場合のうち、各感染世代が  $1 \rightarrow 1 \rightarrow 2$  のような順序の感染者数で与えられるとき、

$$\begin{aligned}
 (2.9) \quad \Pr(1 \rightarrow 1 \rightarrow 2) &= \Pr(I_1 = 1 \mid S_0 = 4, I_0 = 1) \cdot \Pr(I_2 = 2 \mid S_1 = 3, I_1 = 1) \\
 &\quad \cdot \Pr(I_3 = 0 \mid S_2 = 1, I_2 = 2) \\
 &= 12p_1^3 q_1^4 q_2
 \end{aligned}$$

といったように、直接的に手計算することで観察確率が導出できる。上述の通り、Reed-Frost 型モデルは、 $a$  名の感染者がいる集団において 1 人の感受性宿主が接触を回避する確率が  $q_a = q^a$  であると想定する。接触回避確率  $q$  は感染世代に独立であることから、感染世代別の感染確率は該当する世代における感染者数にのみ依存すると想定している (Bailey, 1975)。よって式 (2.9) は  $12p^3 q^6 (= 12p^3 (1-p)^6)$  に単純化される。同様の手続きを他の全てのあり得る流行連鎖について考えることで、多項分布に従う理論的な最終規模分布の期待値が得られる。密な接触が起こりやすい家庭内では、個体別の接触パターンの違いを大まかに無視できる (全員がほぼ同等に密な接触を経験する) と想定していることに注意しよう。家庭内において伝播の異質性を無視できることは家庭内感染のデータがコミュニティのそれと異なる最も大きな特徴の 1 つである。

Reed-Frost 型モデルだけでなく、Greenwood (1931) によって異なる接触回避確率の想定が提案されているので、それも紹介しておこう。Greenwood は英国における麻疹の研究において、「家庭内で少なくとも 1 人の初期感染者が与えられたとき、それによって連鎖的な 2 次感染が起こる確率は各感染世代の感染者数に独立である」と考えた。すなわち、 $a$  名の感染者がいる集団において 1 人の感受性宿主が接触を回避する確率を  $q_a = q$  と想定したのである。このため

Greenwood 型は Reed-Frost 型モデルよりも若干単純な結果を与える(基本構造は大きく異なる). 式(2.9)の結果を Greenwood 型モデルの想定で書き換えれば,  $12p^3q^5(=12p^3(1-p)^5)$  に単純化される. ここで, Reed-Frost 型モデルも Greenwood 型モデルも, 1つのパラメータ  $p$  (1人の感受性宿主が1人の感染者と少なくとも1回の接触を経験する確率)のみを推定することを覚えておこう. 推定される  $p$  は家庭内における感染性の指標として有用と考えられてきた.

図2は1929-34年にロードアイランド州で観察された麻疹の観察データに Greenwood 型モデルを適用した結果である(Wilson et al., 1939). 流行前に全ての家族構成員が麻疹に対して感受性を有すると考えられる3人家族において, うち1人が初期感染者として侵入した場合に, 残りの感受性を有する2人が感染を経験する確率を検討している. よって, 最終規模は1人, 2人および3人という3通りの結果があり, それぞれの予測値とともに上述の議論から得られる多項分布に従う確率を示している. 確率  $p$  の最尤推定値は0.649である.

図3は英国における5人家族664世帯を対象にした, かぜ症候群(common cold)に関する最終規模の観察の一部である(Heasman and Reid, 1961; Schenzle, 1982). Reed-Frost 型モデルおよび Greenwood 型モデルの予測値を比較している. 一般的に家庭の規模が大きくなると, Reed-Frost 型モデルの方がより適切に観察値を描写できる(特に, 長い流行連鎖について Greenwood 型モデルの適合度が悪くなる)傾向が経験的に知られている(Bailey, 1975). 現に, 図3で  $\chi^2$  統計量を用いた適合度検定を実施した場合, Reed-Frost 型および Greenwood 型モデルの  $p$  値はそれぞれ0.059と0.001である. Reed-Frost 型モデルのほうが良いと決定的に判断するものではないが, Greenwood 型モデルが観察値から乖離する傾向を反映している. 一方で, Greenwood 型モデルは世帯の規模が3-4人のように比較的小さく, コミュニティと比較して家庭内で伝播が起こる確率が非常に高い感染症に対して有用であることで知られる. 典型的な疾患例は結核であり, 家庭内における結核の感染予防効果判定に関しては, Greenwood 型モデルが現在も尚, 頻繁に用いられている(Akhtar et al., 2007).

最後に, 図2および図3で用いた期待値の議論を一般化した順列を紹介しておこう. 上述のモデルでは流行の初期感染者1人のみが家庭内に侵入した場合を考えたが, 現実には複数の  $a$  人(仮に  $a \geq 2$  とする)が初期に侵入する場合もあるであろう. まず Greenwood 型モデルを考える. いま, 総数  $n$  人の感受性宿主を有する家庭において(うち初期感染者数が  $a$  人である場合), 最終的に  $j$  人が感染を経験する確率を  ${}_aP_{n,j}$  とする ( $a \leq j \leq n$ ). 感染第1世代において,

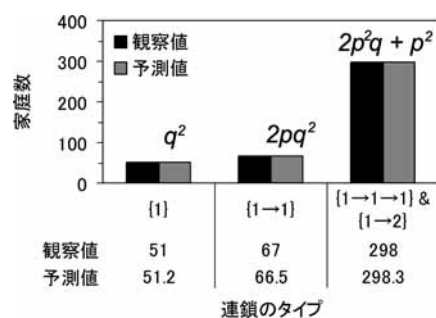


図2. Greenwood 型モデルを利用した家庭内の麻疹感染に関する最終規模. 1929-34年の間にロードアイランド州で観察された麻疹の家庭内感染の観察データに対する適用. 1人の初期感染者と2人の感受性宿主が存在する家庭を対象に最終規模の分布が与えられている(Wilson et al., 1939). 図内  $p$  及び  $q$  はある感染世代において1人の感受性宿主が, ある感染性宿主との間で接触を経験する確率および接触を免れる確率 ( $p = 1 - q$ ) である.

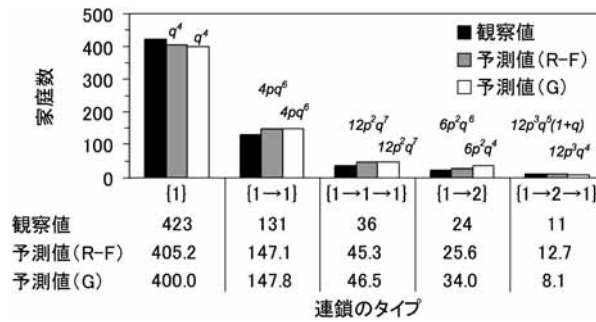


図 3. Reed-Frost 型および Greenwood 型モデルを利用した、かぜ症候群の家庭内感染データの説明. 英国の 5 人家族 664 世帯を対象にしたかぜ症候群の最終規模の観察値とモデル予測値の比較 (Heasman and Reid, 1961; Schenzle, 1982). R-F は Reed-Frost 型モデル, G は Greenwood 型モデル. 図内  $p$  及び  $q$  はある感染世代において 1 人の感受性宿主が, ある感染性宿主との間で接触を経験する確率および接触を免れる確率 ( $p = 1 - q$ ) である (Reed-Frost 型モデルの期待値のほうを常に上に表示している).  $\chi^2$  統計量を利用した適合度検定結果を比較すると, Reed-Frost 型モデルの  $p$  値は 0.059, Greenwood 型モデルの  $p$  値は 0.001 であり, Greenwood 型モデルは長い流行連鎖の想定を必要とする観察データに対応することが難しい点を裏付けている.

$k$  人の感染者が生み出される確率は二項分布を利用して

$$(2.10) \quad \binom{n-a}{k} p^k q^{n-a-k}$$

で与えられる. よって, この世代には  $k$  人の感染者と  $n - a - k$  人の感受性宿主が存在することになる. そのため, 初期感染者  $a$  人を除いて合計で  $j - a$  人の 2 次感染者を生み出す確率は  ${}_k P_{n-a, j-a}$  である. よって,

$$(2.11) \quad {}_a P_{n, j} = \sum_{k=1}^{j-a} \binom{n-a}{k} p^k q^{n-a-k} {}_k P_{n-a, j-a}$$

と書くことができる. 2 次感染者が 1 人も生み出されない確率は,  ${}_a P_{n, a} = q^{n-a}$  である. 同様の手法で Reed-Frost 型モデルの連鎖も一般化すると,

$$(2.12) \quad {}_a P'_{n, j} = \sum_{k=1}^{j-a} \binom{n-a}{k} (1-q^a)^k q^{a(n-a-k)} {}_k P'_{n-a, j-a}$$

であり, 2 次感染者が生み出されない確率は,  ${}_a P'_{n, a} = q^{a(n-a)}$  である (Bailey, 1975). 式 (2.11) と (2.12) があれば, 図 2 および図 3 のような観察データに対して複数の初期感染者に対応した最終規模の尤度方程式 (多項分布に従う観察確率の期待値) を容易に導くことができるだろう.

ここまでに紹介した Reed-Frost 型モデルおよび Greenwood 型モデルは最も古典的の原文のモデル構造を一般化したものであるが, 20 世紀中盤-後半には同様の連鎖型二項分布モデルについて様々な発展と改良が加えられた. 例えば, 家庭内で感受性を有する者のうちで感受性の性質が個体によって異なる場合や (Gart, 1972), 感染個体間で感染性の異質性を考慮した一般的な連鎖型二項分布など (Becker, 1981; Gani and Mansouri, 1987), 現実の観察に可能な限り対応するための工夫が加えられてきた. 本稿では連鎖型二項分布の定量的パフォーマンスに関して述べないが, より詳しい歴史的発展は Dietz and Schenzle (1985) によって解説されており,

さらに連鎖型二項分布の基礎的メカニズムは Bailey (1975) によって初学者向けに詳説されている。Daley and Gani (2001) と Becker (1989) は一般化や詳細な応用的発展に関する議論を展開している。

### 3. 家庭内 2 次発症割合の推定

古典的な連鎖型二項分布モデルは理論的な流行連鎖の特徴をわかりやすく描写しており、かつ尤度方程式の導出や数値計算が容易であることなどの理由から、20 世紀中盤-後半にかけて数多くの拡張と応用が繰り返された。それは、疫学モデルのパラメータ推定を実施する上で、特定の家族構成員数の家庭を対象にした最終規模データが非常に有用であることが広く認識される機会になった。しかし、前節の解説で触れなかった 2 つの問題点を理由に、Reed-Frost 型や Greenwood 型のモデルを全ての感染症の観察データに应用することが不適切な場合も少なくないことが後に明らかにされた (Longini and Koopman, 1982; Longini et al., 1982)。それら理由を端的に書くと、(1) 連鎖型二項分布モデルは初期感染者に条件付けした確率によって描写されるため、コミュニティにおける感染確率を明示的に反映できないこと (式 (2.11) と (2.12) のように初期感染者数を変更する程度しかできない)、および (2) 流行連鎖を考える場合、感染待ち時間 (latent period) が一定の長さである必要があり、かつ、それに引き続いて非常に短い感染性期間の間に 2 次感染が起こることを暗に想定していること、である。後者の問題点は、もし感染性期間が感染待ち時間よりも長い場合 (例えば、インフルエンザやその他のウイルス性のかぜ症候群が該当する)、複数の感染世代における感染者の感染性期間が同時刻で重複する可能性が高く、家庭内の流行連鎖を離散的な感染世代別に明示的に考慮することに限界があることを意味している。これらの問題点を解決する (コミュニティでの感染確率を明示的に推定し、かつ感染世代の重複の問題を排除する) ためには、家庭内の流行連鎖を検討するよりも、無理な想定を施さずに最終規模だけを検討したほうが有用かつ妥当であることも少なくない。同問題の解決手段の 1 つが、提唱者の名前をとって Ludwig 型と呼ばれる確率モデルである (Ludwig, 1970, 1975)。Ludwig 型モデルは、推定されたパラメータを疫学的解釈に関連付けた後の研究の著者名をとって Longini and Koopman (1982) 型と称されることもあるが、モデル構造自体は Ludwig (1970) によって最初に提案されたので本稿では Ludwig 型モデルと呼ぶことにする。その基礎的なモデル構造の考え方をみていこう。

まず、確率  $B$  を 1 人の感受性宿主が 1 つの流行を通じてコミュニティで感染しない確率とし、確率  $Q$  を 1 人の感受性宿主が家庭内にいる 1 人の感染性宿主から感染を免れる確率とする。これを基に最終規模に関する組合せの問題を考えよう。次に述べる家族構成員は流行前に全てが感受性を有するとする。連鎖型二項分布では家族構成員について初期感染者と 2 次感染者を分けて標記したが、実際は多くの観察において感染者が家庭内かコミュニティかいずれで感染したのかが不明であることが多い。よって、ここでは流行前の家族構成員の総数のみが既知とする。家族構成員数が 1 人の家庭において (1 人暮らしの場合)、流行終息後に感染者が 0 人の確率は  $\Pr(0|1) = B$  であり、1 人の感染者を認める確率は  $\Pr(1|1) = 1 - B$  である。家族構成員が 2 人のとき、流行終息後に感染者が 0 人の確率は  $\Pr(0|2) = B^2$  である。また、 $\Pr(1|2)$ 、つまり 2 人のうちどちらか 1 人が感染する確率は 2 通りの組合せがある。つまり、「1 人目が  $1 - B$  の確率でコミュニティにおいて感染するが、2 人目は確率  $B$  でコミュニティから感染を逃れ、かつ 2 人目は確率  $Q$  で家庭内の 1 人目の感染者による 2 次感染から逃れる」という場合、あるいは、これが 1 人目と 2 人目で逆 (1 人目が感染から逃れ、2 人目が感染する) という場合がある。よって  $\Pr(1|2) = 2(1 - B)BQ (= 2\Pr(1|1)BQ)$  である。また、 $\Pr(2|2)$ 、つまり 2 人両方が感染する確率は「両方がコミュニティで感染する」あるいは「一方がコミュニティで感染



し、他方に家庭内で2次感染を引き起こす」であるから(後者は1人目と2人目が入れ替わるので2通りあることに注意しよう),  $\Pr(2|2)=(1-B)^2+2(1-B)BQ(=1-\Pr(0|2)-\Pr(1|2))$ で与えられる.

以上のプロセスを  $s$  人の家族構成員(流行前に  $s$  人の感受性宿主を有する家庭)について一般化しよう. 流行終息後に  $k$  人が感染するという事は,  ${}_sC_k$  通りの感染者の組合せがあることにほかならない.  $k$  人が感染する確率は  $\Pr(k|k)$  である. また,  $s-k$  人は感染しないので, それらの者はコミュニティでの感染を逃れており ( $B^{s-k}$ ), かつ家庭内で  $s-k$  人の全てが  $k$  人の感染者による2次感染から逃れなければならない ( $Q^{k(s-k)}$ ). よって,  $s$  人の家庭で  $k$  人が感染する確率(最終規模を観察する条件付き尤度)は

$$\Pr(k|s) = \binom{s}{k} \Pr(k|k) B^{s-k} Q^{k(s-k)} \quad \text{for } k=0, 1, \dots, s-1;$$

$$\Pr(s|s) = 1 - \sum_{k=0}^{s-1} \Pr(k|s) \quad \text{for } k=s$$

によって与えられる(Ludwig, 1975; Haber et al., 1988).

図4は大阪市における671世帯の3人家族(流行前に3人の感受性宿主を有する家庭)で観察されたアジアかぜ(1957-8年のA型インフルエンザ流行)に関する最終規模である(山本, 1959).  $s=3$ としてモデル(3.1)を適用したものが期待値である. 3人家族の場合, 家庭内感染の最終規模  $k$  は「0人, 1人, 2人あるいは3人全てが感染を経験する」という4通りあり, その離散的な分布に式(3.1)を適用した積をとるだけで確率  $B$  および  $Q$  の最尤推定値が各々0.79および0.88と推定される.

Longini and Koopman (1982)の定義より,  $1-B$ はコミュニティにおける感染確率(**community probability of infection; CPI**)と称され, 感染症伝播におけるコミュニティ関与度を評価する指標として用いられる. 一方で,  $1-Q$ は家庭内における2次感染の確率であり,  $100(1-Q)$ は家庭内で1人の感染者との接触後に発症する者の割合を意味する家庭内2次発症割合(household secondary attack rate)に等しい(感染した者の全てが発症すると仮定した場合).  $Q$ は家庭内の伝播メカニズムを理解する上で示唆に富む. 例えば,  $\beta(\tau)$ を感染齢  $\tau$  (1人の感染個体における感染後の経過時刻を感染齢と呼ぶ)における, 濃密な接触による家庭内の2次感染確率だと

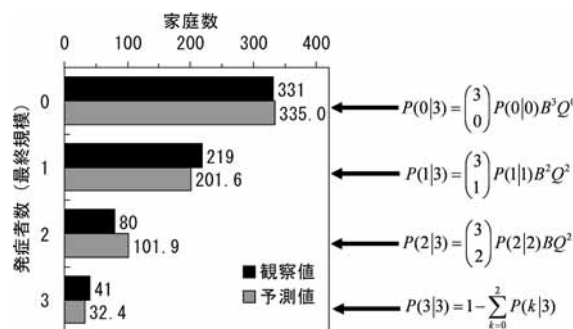


図4. Ludwig型モデルを用いた家庭内2次発症割合およびコミュニティ感染確率の推定. 1958年の大阪市におけるA型インフルエンザ流行(アジアかぜ)の最終規模データ(山本, 1959). 予測値は計671世帯の3人家族を対象にして, Ludwig型モデルを適用したものである. コミュニティにおける感染確率  $(1-B)$  および家庭内2次発症割合  $(100(1-Q))$  の最尤推定値は, それぞれ0.21および12.2%と推定される.

する. さらに  $t_0$ ,  $t_l$  および  $t_m$  をそれぞれ(一感染個体の)感染時刻, 感染性獲得時刻, 感染性消失時刻とする.  $\beta(\tau)$  に以下のような単純な階段関数を仮定しよう:

$$(3.2) \quad \beta(\tau) = \begin{cases} 0, & t_0 \leq \tau < t_l \\ b > 0, & t_l \leq \tau < t_m. \end{cases}$$

例えば, インフルエンザを例に挙げると,  $t_l - t_0$  (感染待ち時間) および  $t_m - t_l$  (感染性期間) はそれぞれ 1.5-2.5 日間および 4 日間程度と想定されることが多い(Longini and Koopman, 1982; Longini et al., 2004). すると,  $Q$  は以下のようにして  $b$  に関連付けることができる:

$$(3.3) \quad Q = \prod_{\tau=t_0}^{t_m} (1 - \beta(\tau)) = (1 - b)^4.$$

つまり, ある特定の家族構成員を有する家庭に関して最終規模の分布を得るだけで, 家庭内伝播の特徴をパラメータ解釈とともに明示的に分析できるのである. 逆に考えれば, 家庭内伝播の最終規模を利用することによって, ある感染症の家庭内における感染性期間を容易に推定することにも役立つだろう(Cauchemez et al., 2004).

モデル(3.1)は極端なケースにおいて他の古典的モデルと容易に関連付けることができる. 仮に家庭内での伝播が全く起こらない場合( $Q=1$ )を考えると, モデル(3.1)は以下の単純な二項分布に置き換えられる(Haber et al., 1988):

$$(3.4) \quad \Pr(k | s) = \binom{s}{k} (1 - B)^k B^{s-k}.$$

また, 家庭内感染のみが起こる極端な場合( $B=1$ )を考えると, モデル(3.1)はReed-Frost型モデルに等しい. つまり, モデル(3.1)は一見すると最終規模のみに基づくように思われるが, その理論的解釈は非常に明示的であり, Reed-Frost型モデルの想定はコミュニティにおける感染を無視することと数理的に等価であると暗示していることを覚えておこう. 同様のモデルは家庭内2次発症割合の推定に非常に有用であるだけでなく, コミュニティにおける感染確率を同時に推定できることから, 現在でも様々な拡張が行なわれている. より一般化されたモデルはAddy et al. (1991)に詳述されているので参考にしていきたい.

Ludwig型モデルの背景にある想定をリストしておこう. 特に(4)番目の想定に注意しよう.

- (1) 対象とする流行において1人の感受性を有する個体は1度だけ感染する可能性がある(複数回の感染は起こらない).
- (2) 全ての個体はある閉鎖的なコミュニティに属する. また, 各個体は1つの家屋に属する. 家屋は1人以上の家族構成員を有する.
- (3) コミュニティにおける感染源は均一に分布している. 家族構成員は家庭内において密で均質な接触機会を有する.
- (4) 各個体は家庭内あるいはコミュニティのいずれかで感染する可能性がある. コミュニティでの感染確率は家庭内での感染確率と独立である.

ちなみに, 本稿の議論は家庭内伝播のみに注目しているが, モデル(3.1)は家庭内とコミュニティだけでなく, その他の感染機会(例えば, 学校や通勤・通学バスなど感染リスクが高いと考えられる場所・機会)を根拠に発展することも可能である(Longini et al., 1988). ただし, その場合も想定(4)を満たす必要があることに注意しよう.

#### 4. 家庭間再生産数を利用した流行閾値

第2節と第3節では、主に家庭内伝播の観察データを利用した推定手法の考え方と推定されるパラメータの疫学的解釈を中心に単純な確率モデルを紹介した。特に、家庭内の最終規模という観察しやすいデータを合理的に説明するメカニズムに焦点を当ててモデル構造を検討したが、それらの構造は単純であることから定量的な応用性に富んだモデルと言えるだろう。また、数値計算をする上において尤度方程式は簡単な手計算で導出できるレベルであり、生物統計学を専門にしない観察者自身がモデルを利用することができるのは大きな利点である。しかし、現実的な家庭内の伝播機構はモデルが想定するよりも遥かに複雑であり、単純なモデルでは本来の複雑な最終規模の分布を記述することに限界を認めることを記憶しておくことが望ましい。最近の位相幾何学の発展に伴う接触ネットワークの研究(各個体を接続する感染のネットワーク構造に関する特徴とその疫学的役割に関する研究)で明らかにされているように、最終規模は伝播の異質性によって大きく変動することが知られている(増田・今野, 2005)。単純な最終規模の分析をするための想定をより深く理解できるように、家庭内伝播に関する別のモデルを検討してみよう。

##### 4.1 家庭内伝播に関するモデルの比較

図5は、家庭内伝播に関するモデルの考え方を比較したものである。最も古典的な Reed-Frost 型モデルおよび Greenwood 型モデルは特定数(多くの場合は1人)の初期感染者に条件付けした家庭内の流行連鎖(感染世代間の伝播)を考慮している。その特徴のために、家庭内の伝播の途中でコミュニティから新たな感染者が侵入する可能性に対応していない。また、前節で議論した通り、現実には異なる感染世代の感染者が同時刻で重複する可能性があり、それは単純な連鎖を考えることの限界を示唆している。Ludwig 型のモデルはコミュニティと家庭内の感染

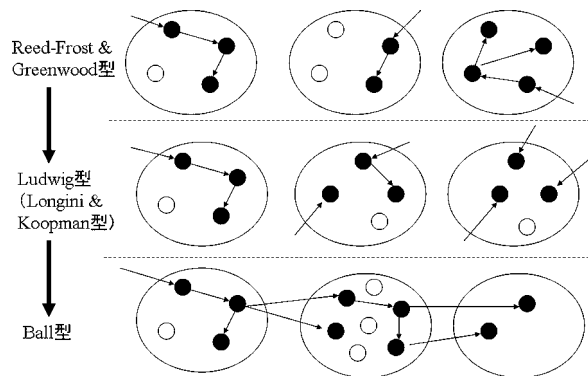


図5. 家庭内伝播に関するモデルの疫学的想定と比較。大きな楕円は家庭 (household) を意味し、小さな円は個体を表す。黒が感染個体であり、白は感染を免れた個体である。矢印は伝播の順序を表す。Reed-Frost 型モデルおよび Greenwood 型モデルは、特定数の初期感染者の侵入に条件付けした流行連鎖を考えており、家庭内伝播の途中でコミュニティから新たな感染者が侵入することに対応していない。Ludwig 型モデル (Longini and Koopman 型モデルとも呼ぶ) は家庭内伝播とコミュニティ伝播を明確に区別して推定できるが、感染者を有するある家庭の観察は他の家庭の観察と独立であることを暗に想定している。Ball 型モデルは Ludwig 型モデルの想定に加えて家庭間の感染の従属性を想定しているが、その流行閾値が集団全体の家庭の世帯人員数や家屋数の分布に依存するために、最終規模のみを基に Ball 型モデルのパラメータを推定することは難しい。

確率を明示的に分けて推定できる上に、連鎖型二項分布モデルの持つ問題点を合理的に解決している。しかし、モデルの構築過程においてコミュニティと家庭内の伝播が独立であることを想定したことに注意しなければならない。すなわち、ある感染者を特定の家屋で観察することは、別の家屋における感染者の観察と独立であることを想定しており、(本稿の導入時に説明した)最も重要な感染症疫学の特徴である非線形の流行動態(動的な新規感染の振る舞い)に対応していない。よって、家庭内の感染確率は(一定の想定を受け入れた上で)感染性期間を推定する目的などに有用かも知れないが、得られた全ての推定値に関する疫学的解釈を与えることは困難であることを示唆している。

ちなみに、コミュニティの感染確率は(想定が妥当と仮定した上で)、最終規模に関する観察情報に加えて集団の総人口  $N$  が明らかな場合に、更なる疫学的解釈を与える。最終規模を  $z(0 \leq z \leq 1)$  とすると、1つの流行における総感染者数は  $zN$  である。ここで、1人の感染者が生み出すコミュニティにおける2次感染者数の平均値(これをコミュニティ再生産数と呼んでおく)が  $R_c$  とすると、コミュニティにおいて感染を引き起こす総接触回数は  $zNR_c$  で与えられる。すると、その集団に属する個体がコミュニティで感染を経験するであろう累積リスクは  $zR_c$  であるから、コミュニティで感染を回避する確率は  $q_e = \exp(-zR_c)$  で与えられる。つまり、 $B$  を利用してコミュニティでの感染回避確率とコミュニティ再生産数との関係を利用した議論を展開することは困難ではない。

ただし、家庭内の伝播頻度が高い感染症を考える場合、コミュニティ再生産数はそれだけで流行閾値を与えるものでなく、家庭内伝播も明示的に考慮しなければならない。コミュニティ再生産数だけを用いて集団全体の最終規模を明示的に検討することは困難であるが、家庭の内外における伝播の異質性を明確に捉えたモデルを利用すれば流行閾値や最終規模の問題を議論することが可能である。それを実現したのが図5のBall型モデルである(Ball and Becker, 2006; Ball and Lyne, 2002a; Ball et al., 1997)。明示的に家庭間の従属性(コミュニティでの接触)を考慮したモデルを理解するために、その構造の基礎と家庭間再生産数の導出方法をみていこう。

#### 4.2 家庭間再生産数

まず、Becker and Dietz (1995)に従って、興味の対象である人口集団を家屋に分けて考える。 $n$  を家庭のサイズ(家族構成員数)とし、 $m_n$  はサイズ  $n$  の世帯数だとする。最大の家族構成員数を  $n_{max}$  とすると、 $m = \sum_{k=1}^{n_{max}} m_k$  および  $N = \sum_{k=1}^{n_{max}} km_k$  は、それぞれ総世帯数および総人口である。この集団において時刻  $t=0$  に初期感染者の侵入が起こった場合を考えよう。議論を単純化するため、流行開始前に全ての個体は感受性を有するとする。感染者は感染性期間を表すランダム変量  $T_I$  に従って分布しているとする。

感染性期間の間、1人の感染個体はコミュニティで感受性個体と接触(global contact)を経験し、感染率  $\lambda_G/N$  のポアソン過程に従って2次感染を起こす。また、同感染個体は家庭内で感受性個体と接触(local contact)を経験し、それは感染率  $\lambda_L$  のポアソン過程に従うものとする。感染を引き起こす接触を説明する全てのポアソン過程および感染性期間を説明するランダム変量は各個体内および個体間で独立であるとする。感受性宿主は感染を経験すると直ちに感染性を獲得するものとする(感染待ち時間を無視する)。つまり、いま興味の対象は流行閾値と最終規模であり、それらは感染待ち時間の長さに独立であることから(Ma and Earn, 2006)、ここでは現実的な感染待ち時間の長さは無視してしまうことにしよう。

流行閾値の問題を議論する場合、集団全体で個体が均質な接触を経験することを想定するモデル(SIR型モデルなど)と同様に、流行初期に感受性宿主の減少を近似的に無視できる期間を対象にした線形モデルを考えるとわかりやすい。つまり、総世帯数  $m$  が十分に大きく、コミュニティ内におけるある接触相手が既に感染者を有する家庭に属する者である確率が十分に

小さいと想定しよう。よって、流行初期にコミュニティ接触の全てが感受性個体のみを有する家庭に属する者に起こることを近似的に想定できるとする。このとき、感染者数の増殖は分岐過程で説明される。ここで線形近似を満たす流行初期において、典型的な感染者を有する家庭から感受性宿主のみを有する家庭へと2次感染が起こる家庭間の伝播数(2次感染家庭数)を  $R$  とする。 $R$  は、コミュニティでの接触回数(従って、家族構成員数の集団内の分布)によって特徴付けられる、子孫の分布(offspring distribution)を与えるランダム変量である。このとき、期待値  $R_* = E(R)$  はモデルの閾値パラメータである。つまり、総世帯数  $m$  が十分に大きいとき、 $R_* > 1$  の場合のみ(超臨界の場合のみ)、非ゼロの確率で大規模流行が生じる。従来の個体群モデルにおける  $R_0$  が(全ての者が感受性を有する集団において)1人の感染者が生み出す2次感染者数の平均値と解釈できるのに対して、 $R_*$  は(全ての家屋が感受性個体のみからなる集団において)1つの感染家庭が生み出す2次感染家庭数の平均値である。これを家庭間再生産数(**household reproduction number**)と呼ぶ。

閾値パラメータ  $R_*$  は以下のようにして導出することができる。対象集団における家族構成員(サイズ)  $n$  の世帯数の総世帯数に対する割合を  $\alpha_n = m_n/m$  とし、コミュニティでの接触が  $n$  人家族に属する1人の個体に起こる確率を  $\tilde{\alpha}_n = nm_n/N$  とする。感受性を有する者のみからなる  $n$  人家族の全てについて、コミュニティにおける接触を完全に無視した場合、 $n$  人家族の家庭内で期待される最終規模の平均値を  $\mu_n(\lambda_L)$  とする。一方で、コミュニティにおいては、1人の感染者が平均  $E(T_I)$  の感染性期間の間に率  $\lambda_G$  で接触を経験することから、1つの感染家庭から生じるコミュニティ接触回数の合計は、 $\mu_n(\lambda_L)\lambda_G E(T_I)$  で与えられる。さらに、それぞれのコミュニティ接触は、対象とする人口においてランダムに選出された個体との間で起こる(以上のことを全ての家庭サイズ  $n$  について考えなければならない)。よって、

$$(4.1) \quad R_* = \sum_{n=1}^{n_{max}} \tilde{\alpha}_n \mu_n(\lambda_L) \lambda_G E(T_I)$$

で家庭間再生産数が与えられる。ここで、 $n$  人家族の家庭内最終規模の平均値を表す  $\mu_n(\lambda_L)$  は、第3節で議論した(コミュニティ感染を無視した場合の)家庭内感染の最終規模と概念的に同じものであり

$$(4.2) \quad \mu_n(\lambda_L) = n - \sum_{k=0}^n \binom{n}{k} \beta_k(\lambda_L) \Phi(\lambda_L k)^{n-k}$$

で与えられる。ここで  $\Phi(\theta) = E(\exp(-\theta T_I))$  は感染性期間  $T_I$  のモーメント母関数であり、 $\beta_k(\lambda_L)$  は以下の解より得られる家庭内感染確率の指標である：

$$(4.3) \quad \sum_{i=0}^k \binom{k}{i} \beta_i(\lambda_L) \Phi(i\lambda_L)^{k-i} = k.$$

本稿の第2節と第3節で最も基本的なモデルを既に議論したので、それよりも複雑な式(4.2)と(4.3)の導出方法を省略するが、これらの詳細は Ball (1986) および Ball et al. (1997) を参考にしたい。

ここで、式(4.1)で与えられる家庭間再生産数  $R_*$  と個体レベルで定義される基本再生産数  $R_0$  の関係を考えよう。式(4.1)から明らかな通り、 $R_*$  は集団内の家屋数や家族構成員数の分布に影響を受けるため、それらの現実的な分布を利用するときは  $R_*$  と  $R_0$  の関係は数値的にしか比較することができない。しかし、極端な例において、集団が理論的に単一サイズの家族構成員を有する家庭のみで構成される場合、 $R_0$  と  $R_*$  を解析的に比較することが可能である。最も単純な例として、集団における家族構成を全て無視できる場合を考えてみよう。つまり、全て

の家屋における家族構成員数が  $n=1$  と想定することによって、全ての個体が均質な接触を経験するモデルにおける  $R_0$  と  $R_*$  の関係を解析的に検討することができる。このとき

$$(4.4) \quad R_0 = \lambda_G E(T_I)$$

であることは明らかだから (Heesterbeek and Dietz, 1996),

$$(4.5) \quad \mu = \sum_{n=1}^{n_{max}} \tilde{\alpha}_n \mu_n(\lambda_L)$$

とすれば  $R_* = R_0 \mu$  である。  $\mu$  は家庭内伝播の影響に伴う閾値パラメータの増幅係数と解釈することができるだろう。ここで、式(4.4)の  $R_0$  の解釈は「全ての家庭がサイズ1のとき」の  $R_0$  として単純な関係が与えられていることに注意しよう。同様に、集団の全てが  $n$  人家族で構成されるとすると、その集団全体で家庭内の接触は  $(n-1)\lambda_L$  の率で起こり、コミュニティの接触は  $\lambda_G$  の率で独立に起こる。これらが平均  $E(T_I)$  の感染性期間だけ続くのだから、「全ての家庭がサイズ  $n$  のとき」の  $R_0$  は

$$(4.6) \quad R_0 = (\lambda_G + (n-1)\lambda_L)E(T_I)$$

で与えられる。ここで、 $\lambda'_L = (n-1)\lambda_L$  とし、感染性期間  $T_I$  は定数で1日(よって  $\Phi(\theta) = \exp(-\theta)$  を得る)と想定すると、 $R_0$  と  $R_*$  の関係を  $\lambda'_L$  と  $\lambda_G$  の関数として感度分析することができる。家族構成員  $n \geq 2$  の人口において、 $\lambda_L$  が十分に小さい範囲では臨界的なレベル  $R_* = 1$  が  $R_0 = 1$  とほぼ一致する (Ball and Lyne, 2006; Ball et al., 1997; Becker and Dietz, 1995)。しかし、式(4.4)の極端な例からわかるように、家族構成員数  $n$  が極端に小さい家庭ばかりで構成される集団では、 $R_0 > 1$  と  $R_* > 1$  は一致しないことを覚えておこう。  $R_0$  と  $R_*$  の関係だけでなく、流行時刻  $t$  の関数で与えられる閾値パラメータの1つである実効再生産数 (effective reproduction number;  $R(t)$ ) を考える場合も、同様の方法で  $R(t)$  と  $R_*(t)$  の関係が得られることが知られている (Fraser, 2007)。

一方で、このモデルにおける最終規模はどうだろうか。  $R_0$  と同様に、最終規模も家族構成員や家屋数の分布に影響を受けるため、現実の家族構成員や家屋数に関する統計データを利用する場合は最終規模の分布を解析的に明らかにすることが難しい。しかし、第4.1節でコミュニティ再生産数について議論したように、最終規模  $z$  に関する近似的な議論は可能である。ここで  $K_n$  をサイズ  $n$  の家庭で最終的に感染を経験する者の数を表すランダム変量とする。よって、流行終息後にサイズ  $n$  の家庭からランダムにある個体を選んだとき、選ばれた個体が感染を経験した者である確率は  $E(K_n)/n$  である。ここで、 $E(K_n)$  を議論するために流行の非線形性を一時的に無視する。つまり、流行を通じて単位時間当たりに感染者が存在する確率は人口全体の最終規模  $z$  と総人口  $N$  および平均感染性期間を利用して  $zNE(T_I)$  であると近似する(感染性宿主の有病割合を時刻に対して平均することと同じ手続きをしている)。よって、ある個体がコミュニティ感染を免れる確率  $\pi$  は

$$(4.7) \quad \pi = \exp\left(-\frac{\lambda_G}{N} zNE(T_I)\right) = \exp(-\lambda_G zE(T_I))$$

である。異なる複数の個体がコミュニティ感染を逃れる確率は近似的に独立であることから、 $K_n$  は1つの家屋における流行の最終規模分布に近似される。つまり、

$$(4.8) \quad E(K_n) = \sum_{k=1}^n \binom{n}{k} (1-\pi)^k \pi^{n-k} \mu_{n-k,k}$$

である。ここで  $\mu_{n,a}$  とは Reed-Frost 型モデルの一般化で議論したときと同様に、 $a$  人の初期感染者が  $n$  人家族に侵入したときの家庭内の最終規模の平均値である。よって、興味の対象とする集団全体の最終規模  $z$  は以下を満たす：

$$(4.9) \quad z = \sum_{n=1}^{\infty} \tilde{\alpha}_n n^{-1} \sum_{k=1}^n \binom{n}{k} (1-\pi)^k \pi^{n-k} \mu_{n-k,k}.$$

これは非線形モデルから与えられる明示的な解ではないが、式(4.7)と(4.9)は最終規模  $z$  に関して近似的な示唆を与える (Ball and Lyne, 2002a; Ball et al., 1997)。ちなみに、上述した  $R_*$  と  $R_0$  の関係に関する議論と同様、ここで対象集団における全ての家庭が家族構成員 1 人のみで構成される極端な例を考える。すると、式(4.7)と(4.9)は以下に単純化される：

$$(4.10) \quad \begin{aligned} z &= 1 - \exp(-\lambda_G E(T_I) z) \\ &= 1 - \exp(-z R_0). \end{aligned}$$

これは集団内の全ての個体が均質に接触することを想定したときの最終規模方程式 (final size equation; 1 つの流行における最終規模と伝播能力との関係を表す方程式) に一致する (Ma and Earn, 2006)。

最後に、家庭間再生産数を利用した効果的なワクチン接種戦略の基礎的な考え方についてみてみよう。Ball 型モデルと称される家庭内・家庭間伝播モデルは家庭間再生産数の導出や最終規模の複雑な問題の議論だけでなく、閾値定理 (再生産数が 1 を上回るか否かによって大規模流行の確率が非ゼロか否かを決定する定理) を利用した効果的なワクチン接種戦略の提案に有用であることがよく知られている (Ball and Becker, 2006; Ball and Lyne, 2002b, 2006; Ball et al., 2004a, 2004b; Becker and Hall, 1996)。まず式(4.1)を以下のように書き換える：

$$(4.11) \quad \begin{aligned} R_* &= \lambda_G E(T_I) \sum_{n=1}^{\infty} \tilde{\alpha}_n \mu_n \\ &= \lambda_G E(T_I) \sum_{n=1}^{\infty} \frac{nm_n}{N} \mu_n \\ &= \frac{\lambda_G E(T_I)}{N} \sum_{k=1}^m n_k \mu_{n_k}. \end{aligned}$$

ここで世帯数  $m$  は  $1, 2, \dots, m$  とラベルしているが、これは対応する家族構成員数  $n_1, n_2, \dots, n_m$  に対応している。ここでワクチン接種によって、感受性宿主の数が  $n_k$  人から  $n'_k$  人 ( $k = 1, 2, \dots, m$ ) に減少したときを考えよう。つまり、集団内のワクチン接種者の総数  $N_v$  が

$$(4.12) \quad N_v = \sum_{k=1}^m (n_k - n'_k)$$

の場合を考える。このとき、ワクチン接種を実施した集団における閾値パラメータ  $R_v$  は

$$(4.13) \quad R_v = \frac{\lambda_G E(T_I)}{N} \sum_{k=1}^m n'_k \mu_{n'_k}$$

で与えられることは式(4.11)より明らかである (Ball and Lyne, 2002a, 2002b)。つまり、如何に効果的なワクチン接種戦略を行なうかは、 $R_v$  を 1 未満にする (あるいは少なくとも  $R_v$  を最小にする)  $n'_k$  の分配手段を考察することによって議論することができる。実際にはワクチン接種の生物学的効果は個体の感受性を低下させるだけでなく、仮に感染しようともワクチン接種が感染性を減弱させることもあり (Becker et al., 2003, 2006)、更なる議論の発展は未だ開発途上

の段階にある。具体的な研究例として、Ball and Becker (2006) および Ball and Lyne (2002b, 2006) が詳しいので参考にさせていただきたい。

## 5. 考察

本稿は家庭内伝播に関する基礎的な考え方を養うために、歴史的発展から最近の研究に至るまで、関連するモデルの構造と観察データとの関係を明らかにしながら、家庭内伝播の確率モデルの基礎について簡潔に解説した。特定の家族構成員の家庭で観察された最終規模を明示的に分析することが可能なモデルについては、歴史的な疫学データを分析しつつ議論したが、本稿では推定値そのものよりもモデル自体の考え方とモデル構造の違いを明らかにすることに焦点を絞って議論した。連鎖型二項分布モデルは 1 人あるいは複数の初期感染者が侵入した家庭内における流行連鎖(家庭内における個体間の伝播)を感染世代毎に考えるモデルであり、多項分布の期待値と尤度方程式の導出が容易である。Reed-Frost 型および Greenwood 型モデルの基礎的な仕組みを議論したが、これらのモデルは初期感染者の侵入後にコミュニティにおける感染者の更なる侵入に対応していないことや、感染待ち時間に対して感染性期間が十分に長い場合に異なる感染世代の感染性宿主が同時刻に重複して存在することなどの問題点を有する。Ludwig 型モデルは最終規模の分布のみを描写する確率モデルであり、コミュニティ感染確率と家庭内 2 次発症割合を同時に推定することができる上に、連鎖型二項分布モデルの有する問題点を合理的に解決したものである。しかし、同モデルはコミュニティ感染確率と家庭内感染確率が独立であると暗に想定しており、言い換えるならば(コミュニティの感染確率を議論しているにも関わらず)1 つの家庭における観察は他の家庭における観察と独立であることを想定していることを議論した。よって、コミュニティ感染確率が高い感染症を扱う場合、異なる家庭間の従属性を明示的に考慮した Ball 型モデルを利用することが望ましい。Ball 型モデルの基本構造は分岐過程によって描写され、家庭間再生産数と効果的なワクチン接種戦略を議論するのに役立つ。ただし、特殊な家族構成員数の分布を除いて、基本再生産数や集団全体の最終規模を解析的に明らかにすることが難しいことを議論した。

この総説を通じて読者の方々に伝えたいことは、家庭内伝播に関する考え方とデータの見方の理論的基礎である。本稿は「どのモデルがどの感染症に最も有用か」ということを議論したわけではなく、また、「どの感染症のコミュニティ感染確率が高いのか」を定量的に明らかにしたわけでもない。それよりも、モデルの背景にある想定と構造自体の考え方について、歴史的発展を踏まえて網羅的に仕組みをシェアしていただきたいと感じたことが本稿執筆の動機である。よって、例えば、今日では既に Reed-Frost 型モデルは有用でない、と主張しているわけではないことに注意していただきたい。現に、Greenwood 型モデルも含めて連鎖型二項分布は未だ多くの事例で応用されているし、さらには家庭内伝播だけでなく一般的な流行に対する反復方程式の新たな発展も数多く提案されている(Daley and Gani, 2001; Ferrari et al., 2005)。また、Ludwig 型モデルが連鎖型二項分布モデルよりも必ずしも有用なわけではないことにも注意していただきたい。推定されるパラメータの疫学的意味合いも異なることから、「連鎖型二項分布モデルの想定が妥当でない場合に Ludwig 型モデルを用いる必要がある」というような思考によって応用事例を考えることも不適切である。また、Ball 型モデルによって最終規模を数値的に分析することが最良の結果に繋がるわけではないことにも注意しよう。そういった妥当性に関する細かな議論よりも、むしろ常に「最も現実を合理的に捉えることができる手段は何か」と「いま扱っているモデルは目的とする疫学的情報の推定値を与えるか」を吟味することのほうが重要である。そのためには、観察データを扱う際には必ず各モデルの想定を吟味する必要がある、モデルが現実的な感染現象を捉えているかを判断する材料の 1 つとしてモデルの



構造と特徴を理解しておくことが欠かせない。本稿がこの点を改善するための基礎資料となれば幸甚である。

感染症の数理モデルには「ベストの手法」がない。余りにも単純で均質な接触を想定する個体群動態モデルは現実的な感染症伝播を反映しないが、豊富な解析的情報を与える点で流行メカニズムを理解する示唆に富む。一方で、様々な現実性を詳細に反映した大規模シミュレーションは具体的な感染症の流行を数値的に分析することに有用ではあるが、一方で解析的な示唆は得られないし、定性的な流行の振る舞いを理解せずしてシミュレーションを実施すると不適切な結論を導いてしまう危険性もないわけではない。データの種類と目的に応じて臨機応変に技術を使い分けなければならないのが現状であるが、そのためにはモデル構造と理論的特徴を理解することに加えて「感染症がどのような振る舞いによって流行するのか」を見る目を養わなければならない。この点で、感染症の家庭内伝播は、実際の流行において伝播の異質性を特徴付ける最も重要な一側面に着眼したテーマであり、数値的だけでなく解析的にも明示的に分析することが可能な課題である。家庭内伝播のモデルには未だ解決していない問題も多く、豊富な発展性と様々な応用の示唆を与える。

#### 参 考 文 献

- Abbey, H. (1952). An examination of the Reed-Frost theory of epidemics, *Human Biology*, **24**, 201–233.
- Addy, C. L., Longini, I. M. and Haber, M. (1991). A generalized stochastic model for the analysis of infectious disease final size data, *Biometrics*, **47**, 961–974.
- Akhtar, S., Carpenter, T. E. and Rathi, S. K. (2007). A chain-binomial model for intra-household spread of Mycobacterium tuberculosis in a low socio-economic setting in Pakistan, *Epidemiology and Infection*, **135**, 27–33.
- Andersson, H. and Britton, T. (2000). *Stochastic Epidemic Models and Their Statistical Analysis*, Lecture Notes in Statistics, No. 151, Springer-Verlag, New York.
- Bailey, N. T. J. (1964). *The Elements of Stochastic Processes with Applications to the Natural Sciences*, Wiley, New York.
- Bailey, N. T. J. (1975). *The Mathematical Theory of Infectious Diseases and Its Applications*, 2nd ed., Charles Griffin and Company Ltd., London.
- Ball, F. G. (1986). A unified approach to the distribution of total size and total area under the trajectory of infectives in epidemic models, *Advances in Applied Probability*, **18**, 289–310.
- Ball, F. and Becker, N. (2006). Control of transmission with two types of infection, *Mathematical Biosciences*, **200**, 170–187.
- Ball, F. and Lyne, O. (2002a). Epidemics among a population of households, *Mathematical Approaches for Emerging and Reemerging Infectious Diseases. Part II* (eds. C. Castillo-Chavez, S. Blower, P. van den Driessche, D. Kirschner and A. A. Yakubu), 115–142, IMA Volumes in Mathematics and Its Application, Springer-Verlag, New York.
- Ball, F. and Lyne, O. (2002b). Optimal vaccination policies for stochastic epidemics among a population of households, *Mathematical Biosciences*, **177/178**, 333–354.
- Ball, F. and Lyne, O. (2006). Optimal vaccination schemes for epidemics among a population of households, with application to variola minor in Brazil, *Statistical Methods in Medical Research*, **15**, 481–497.
- Ball, F., Mollison, D. and Scalia-Tomba, G. (1997). Epidemics with two levels of mixing, *Annals of Applied Probability*, **7**, 46–89.

- Ball, F., Britton, T. and Lyne, O. (2004a). Stochastic multitype epidemics in a community of households: Estimation and form of optimal vaccination strategies, *Mathematical Biosciences*, **191**, 19–40.
- Ball, F., Britton, T. and Lyne, O. (2004b). Stochastic multitype epidemics in a community of households: Estimation of threshold parameter  $R_*$  and secure vaccination coverage, *Biometrika*, **91**, 345–362.
- Becker, N. G. (1981). A general chain binomial model for infectious diseases, *Biometrics*, **37**, 251–258.
- Becker, N. G. (1989). *Analysis of Infectious Disease Data*, Chapman and Hall, New York.
- Becker, N. G. and Dietz, K. (1995). The effects of household distribution on transmission and control of highly infectious diseases, *Mathematical Biosciences*, **127**, 207–219.
- Becker, N. G. and Hall, R. (1996). Immunization levels of preventing epidemics in a community of households made up of individuals of various types, *Mathematical Biosciences*, **132**, 205–216.
- Becker, N. G., Britton, T. and O'Neill, P. D. (2003). Estimating vaccine effects on transmission of infection from household outbreak data, *Biometrics*, **59**, 467–475.
- Becker, N. G., Britton, T. and O'Neill, P. D. (2006). Estimating vaccine effects from studies of outbreaks in household pairs, *Statistics in Medicine*, **25**, 1079–1093.
- Cauchemez, S., Carrat, F., Viboud, C., Valleron, A. J. and Boelle, P. Y. (2004). A Bayesian MCMC approach to study transmission of influenza: Application to household longitudinal data, *Statistics in Medicine*, **23**, 3469–3487.
- Daley, D. J. and Gani, J. (2001). *Epidemic Modelling*, Cambridge University Press, Cambridge.
- Daniel, T. M. (2004). *Wade Hampton Frost, Pioneer Epidemiologist 1880–1938: Up to the Mountain*, University of Rochester Press, New York.
- Dietz, K. (1988). The first epidemic model: A historical note on P. D. En'ko, *Australian Journal of Statistics Series A*, **30**, 56–65.
- Dietz, K. and Heesterbeek, J. A. (2002). Daniel Bernoulli's epidemiological model revisited, *Mathematical Biosciences*, **180**, 1–21.
- Dietz, K. and Schenzle, D. (1985). Mathematical models for infectious disease statistics, *A Celebration of Statistics* (eds. A. C. Atkinson and S. E. Fienberg), 167–204, Springer-Verlag, New York.
- En'ko, P. D. (1989). On the course of epidemics of some infectious diseases, *International Journal of Epidemiology*, **18**, 749–755 (the original publication in Russian in 1889 was translated to English by Klaus Dietz).
- Ferrari, M. J., Bjornstad, O. N. and Dobson, A. P. (2005). Estimation and inference of  $R_0$  of an infectious pathogen by a removal method, *Mathematical Biosciences*, **198**, 14–26.
- Fine, P. E. (1977). A commentary on the mechanical analogue to the Reed-Frost epidemic model, *American Journal of Epidemiology*, **106**, 87–100.
- Fraser, C. (2007). Estimating individual and household reproduction numbers in an emerging epidemic, *PLoS ONE*, **8**, e758.
- 藤曲哲郎 (2003). 『確率過程と数理生態学』, 日本評論社, 東京.
- Gani, J. and Mansouri, H. (1987). Fitting chain binomial models to the common cold, *Mathematical Scientist*, **12**, 31–37.
- Gart, J. J. (1972). The statistical analysis of chain-binomial epidemic models with several kinds of susceptibles, *Biometrics*, **28**, 928–930.
- Greenwood, M. (1931). On the statistical measure of infectiousness, *Journal of Hygiene, Cambridge*, **31**, 336–351.
- Haber, M., Longini, I. M. and Cotsonis, G. A. (1988). Models for the statistical analysis of infectious disease data, *Biometrics*, **44**, 163–173.

- Heasman, M. A. and Reid, D. D. (1961). Theory and observation in family epidemics of the common cold, *British Journal of Preventive and Social Medicine*, **15**, 12–16.
- Heesterbeek, J. A. P. and Dietz, K. (1996). The concept of  $R_0$  in epidemic theory, *Statistica Neerlandica*, **50**, 89–110.
- 平山 雄(1958). 『疫学』, 續文堂, 東京.
- 金光正次, 岡田 博, 甲野礼作, 重松逸造, 平山 雄(1966). 『疫学とその応用』, 南山堂, 東京.
- Karlin, S. and Taylor, H. M. (1975). *A First Course in Stochastic Processes*, 2nd ed., Academic Press, New York.
- Koopman, J. S. and Longini, I. M. (1994). The ecological effects of individual exposures and nonlinear disease dynamics in populations, *American Journal of Public Health*, **84**, 836–842.
- Longini, I. M. and Koopman, J. S. (1982). Household and community transmission parameters from final distributions of infections in households, *Biometrics*, **38**, 115–126.
- Longini, I. M., Koopman, J. S., Monto, A. S. and Fox, J. P. (1982). Estimating household and community transmission parameters for influenza, *American Journal of Epidemiology*, **115**, 736–751.
- Longini, I. M., Koopman, J. S., Haber, M. and Cotsonis, G. A. (1988). Statistical inference for infectious diseases. Risk-specific household and community transmission parameters, *American Journal of Epidemiology*, **128**, 845–859.
- Longini, I. M., Halloran, M. E., Nizam, A. and Yang, Y. (2004). Containing pandemic influenza with antiviral agents, *American Journal of Epidemiology*, **159**, 623–633.
- Ludwig, D. (1970). *Stochastic Population Theories*, Lecture Notes in Biomathematics, Vol. 3, Springer-Verlag, New York.
- Ludwig, D. (1975). Final size distributions for epidemics, *Mathematical Biosciences*, **23**, 33–46.
- Ma, J. and Earn, D. J. (2006). Generality of the final size formula for an epidemic of a newly invading infectious disease, *Bulletin of Mathematical Biology*, **68**, 679–702.
- 増田直紀, 今野紀雄(2005). 『複雑ネットワークの科学』, 産業図書, 東京.
- Maxcy, K. F. (1941). *Papers of Wade Hampton Frost*, Commonwealth Fund and Oxford University Press, New York.
- 西浦 博(2008). 感染症数理モデルのデータサイエンス, 『感染症の数理モデル』(稲葉 寿 編), 60–111, 培風館, 東京.
- 西浦 博, 稲葉 寿(2006). 感染症流行の予測: 感染症数理モデルにおける定量的課題, *統計数理*, **54**, 461–480.
- Nishiura, H., Dietz, K. and Eichner, M. (2006). The earliest notes on the reproduction number in relation to herd immunity: Theophil Lotz and smallpox vaccination, *Journal of Theoretical Biology*, **241**, 964–967.
- 阪本州弘(1985). 『疫学と疫学モデル』, 金芳堂, 京都.
- Schenzle, D. (1982). Problems in drawing epidemiological inferences by fitting epidemic chain models to lumped data, *Biometrics*, **38**, 843–847.
- Tsutsui, T., Minami, N., Koiwai, M., Hamaoka, T., Yamane, I. and Shimura, K. (2003). A stochastic-modeling evaluation of the foot-and-mouth-disease survey conducted after the outbreak in Miyazaki, Japan in 2000, *Preventive Veterinary Medicine*, **61**, 45–58.
- Wilson, E. B., Bennett, C., Allen, M. and Worcester, J. (1939). Measles and scarlet fever in Providence, R. I., 1929–1934 with respect to age and size of family, *Proceedings of the American Philosophical Society*, **80**, 357–476.
- 山本皓一(1959). インフルエンザの家庭内流行様式に関する理論疫学的研究, *大阪市立大学医学雑誌*, **9**, 235–246.

## Stochastic Models for Analysis of Household Transmission Data: Examining Human-made Transmission Experiments

Hiroshi Nishiura

Theoretical Epidemiology, University of Utrecht

To date, various epidemiologic models have been proposed for analyzing infectious disease data, most notably, based on branching processes, birth-death processes and spatially-structured contact processes. To appropriately quantify the transmission dynamics using the observed data, the model has to capture crucial elements of the heterogeneous patterns of transmission. Household transmission plays a key role in characterizing the most important aspects of heterogeneities for many directly transmitted diseases. Household transmission can be deemed a human-made experiment where all family members experience close contact. This article reviews essential pre-requisites of the data analysis and several applications of household transmission models to the data, specifying each of the epidemiologic assumptions and offering detailed insights into the actual mechanisms of household transmission. In particular, it focuses on two topics: a probability model for the analysis of final size distribution in households and another transmission model for deriving the household reproduction number, a threshold parameter of an epidemic. Whenever a household transmission model is employed to estimate key epidemiologic parameters, it is critical to assess the validity and reality of underlying assumptions with respect to disease and data.