

潜伏期間に対数正規分布を仮定した 集団食中毒の曝露時点の最尤推定法

丹後 俊郎*

本論文では、腸管出血性（病原性）大腸菌 O-157:H7 に代表される集団食中毒の曝露時点の推定方法を新しく提案する。感染食品に一齐曝露を仮定し、潜伏期間に対数正規分布を仮定するもので、曝露時点の最尤推定量の簡単な計算方法を提案するとともに、その信頼区間の計算法も提案する。また、従来から文献によく登場している3つの方法（Sartwellの方法、平山の方法、堀内らの方法）は、理論上、または、計算上、妥当性を欠くものであることを示す。本方法の適用例として腸管出血性（病原性）大腸菌 O-157:H7 による集団食中毒事例3件、O-118:H2による事例1件のデータを解析する。

Key words : 腸管出血性大腸菌, O-157:H7, 最尤推定量, 信頼区分, プロファイル尤度

I 緒 言

平成8年、大阪で勃発した腸管出血性大腸菌 O-157:H7 による食中毒の集団発生は食中毒の恐ろしさを再認識させるとともに、当時の菅厚生大臣の「貝割れ大根が感染原因でないことが否定できない」旨の発言によるカイワレ・パニックは食中毒の感染原因特定の重要性とその困難性を浮き彫りにした。

感染原因特定のための重要な疫学的アプローチの一つとして、患者発生時間分布を利用して感染源に曝露した曝露時点の推定方法が考えられる。しかし、曝露時点の推定の方法論に関しては、著者の知る限り最近の研究報告はなく、文献的にはいずれもコンピュータが普及していない1960年代もしくはそれ以前に提案された方法が疫学専門書で紹介されているにすぎない。しかも、過去に提案され、専門書で紹介されている方法の多くは、いずれも一齐曝露でかつ潜伏期間に対数正規分布を仮定するものであるが、驚くことに「理論的に不適切、または、推定精度の悪い」ものであった。

本論文でも、一齐曝露と潜伏期間に対数正規分布を仮定するが、曝露時点に関して漸近的に最良

の推定値が得られる最尤推定法の簡単な計算法を提案するとともに、プロファイル尤度法による信頼区間の推定法も提案する。本方法の具体的な適用例として、腸管出血性大腸菌 O-157:H7 による集団食中毒事例3件、O-118:H2による事例1件のデータを解析する。

II 従来の方法の問題点

文献によく登場する方法は、Sartwellの方法¹⁾、平山の方法²⁾と堀内らの方法³⁾の3種類であるが、Sartwellの方法と平山の方法は基本的に同じである。いずれも、一齐曝露、潜伏期間に対数正規分布を仮定した方法である。曝露時点 γ として、症状が現れる時点 x とすると、 x の確率密度関数は $x > \gamma$ の場合、

$$f(x; \gamma, \mu, \sigma^2) = \frac{1}{\sigma(x-\gamma)\sqrt{2\pi}} \times \exp\left[-\frac{1}{2}\left(\frac{\ln(x-\gamma)-\mu}{\sigma}\right)^2\right] \quad (1)$$

で与えられる。ここに μ, σ はそれぞれ、潜伏期間 $(x-\gamma)$ の対数変換値 $\ln(x-\gamma)$ の期待値(母平均)と母標準偏差である。

1. Sartwell¹⁾の方法

患者発生時点のデータから3つのパーセント点、 $b_1=100p\%$ 点、 $b_0=50\%$ 点、 $b_2=100(1-p)\%$ 点を読みとる(p の値は適当にとる)。すると、曝露時点 γ は対数正規分布の性質より

* 国立公衆衛生院疫学部理論疫学室
連絡先: 〒108-8638 東京都港区白金台4-6-1
国立公衆衛生院疫学部理論疫学室 丹後俊郎

$$\hat{\gamma} = \hat{\gamma}_S(p) = \frac{b_1 b_2 - b_0^2}{b_1 + b_2 - 2b_0} \quad (2)$$

と推定する。しかし、この「関係式」を推定値として用いることは妥当ではない。適当に選んだパーセント点 b_1, b_2 に強く依存して大きく変動し、統計学的性質は極めて悪いからである。Meynell and Williams⁴⁾も同様の指摘をしている。

2. 平山²⁾の方法

平山のテキスト²⁾、金光他のテキスト⁵⁾で紹介されている方法、中村⁶⁾で紹介されている平山の方法は本質的には Sartwell の方法である。しかし、金光他のテキストにある対数確率紙の利用に基づく方法も平山の方法として知られているようである。その、計算手順は次のとおり：

- (a) 仮の曝露時点 s を原点として患者発生時点までの時間の累積分布を対数正規確率紙上にプロットし、ほぼ直線上に並んだ点の中心に合わせて直線を引く。
- (b) 確率紙上に引いた直線を利用して Sartwell と同様に3つのパーセント点、例えば、 $b_1=16\%$ 点、 $b_0=50\%$ 点、 $b_2=84\%$ 点を読みとる。
- (c) Sartwell の式(2)を利用して曝露時点計算する。

この方法は、仮の原点 s が適切であれば、式(2)で計算される曝露時点 γ は仮の原点 s と一致しなければならない ($\gamma=s$) という意味で滑稽な方法で、逆に、仮の原点 s が不適切であれば式(2)で何を推定しているのかが分からないという意味で誤りである。

3. 堀内ら³⁾の方法

文献的には定差図法と呼ばれており、この方法を紹介しているテキストが少なくない⁷⁻⁹⁾。対数正規分布の性質 $\ln(x-\gamma) = \sigma t + \mu$ (t : 正規偏差) より、 x を t の関数として $x(t)$ で表現すれば、

$$x(t) = e^{\mu + \sigma t} + \gamma$$

となる。また、任意の h に対して

$$x(t+h) = e^{\mu + \sigma(t+h)} + \gamma = e^{\sigma h} x(t) + \gamma(1 - e^{\sigma h})$$

となるので、 γ は、 $x(t)$ を x 座標に、 $x(t+h)$ を y 座標にしてプロットして近似された直線と $y=x$ との交点の x 座標として推定する方法である。しかし、この方法では、適当に決めた h に基づく $x(t)$ の曲線の決定が容易でなく、またその推定誤差が γ の推定に大きな影響を与えることは明瞭である。したがって、この職人芸的な方法

は、計算機の普及していなかった時代ならともかく、今日の計算機環境と統計学の学問水準からいっても推薦できる方法ではない。

III 提案する最尤推定法

前節にみたように、これまで提案されている方法はいずれも統計学的にみて明らかに妥当性を欠く方法である。公衆衛生学、疫学の分野の調査・解析の方法論は統計学に基礎を置いているはずである。30年以上も前に提案され、しかも、適切とは言いがたい方法が、なぜこれまでに指摘され改善されてこなかったのだろうか？

問題は3パラメータ対数正規分布の原点パラメータ γ の推定問題に他ならない。このような推定問題には、標本サイズが小さくない限り、漸近的に最良である最尤推定量が利用できることは統計学ではよく知られている¹⁰⁻¹²⁾。したがって、ここでも最尤推定法の利用を提案するが、従来から提案されている最尤推定量の計算方法にくらべて簡単な方法を提案する。

まず、 n 例の食中毒患者の症状の発生時点 $\{x_i, i=1, \dots, n\}$ のデータから計算される尤度関数は、一斉曝露で独立に発生するという条件の下で

$$\begin{aligned} L(\gamma, \mu, \sigma^2) &= \prod_{i=1}^n f(x_i; \gamma, \mu, \sigma^2) \\ &= (2\pi\sigma^2)^{-n/2} \prod_{i=1}^n (x_i - \gamma)^{-1} \\ &\quad \times \exp \left[-\frac{1}{2} \sum_{i=1}^n \left(\frac{\ln(x_i - \gamma) - \mu}{\sigma} \right)^2 \right] \end{aligned}$$

となる。3つのパラメータ (γ, μ, σ^2) の最尤推定量は、対数尤度関数の偏微分を計算して、連立方程式 $(\partial \ln L / \partial \gamma) = (\partial \ln L / \partial \mu) = (\partial \ln L / \partial \sigma^2) = 0$ の解としてもとめるのが通常であるが、この非線形方程式には解が収束しないケースが少なくないという計算上の問題点が知られている^{12,13)}。そのためいろいろな工夫がされているが、ここでは、簡単でかつ収束問題のない線形探索法を利用して解を計算する方法を提案する。

まず、 γ を所与とすると、 (μ, σ^2) の最尤推定量は簡単に

$$\hat{\mu} = \hat{\mu}(\gamma) = \frac{1}{n} \sum_{i=1}^n \ln(x_i - \gamma) \quad (3)$$

$$\hat{\sigma}^2 = \hat{\sigma}^2(\gamma) = \frac{1}{n} \sum_{i=1}^n \{ \ln(x_i - \gamma) - \hat{\mu}(\gamma) \}^2 \quad (4)$$

と計算できる。したがって、最大対数尤度は γ の関数として

$$\ln L^{**}(\gamma) = -n(\hat{\mu}(\gamma) + \ln \hat{\sigma}(\gamma)) \quad (5)$$

と計算できる(ただし、定数項は省略した)。この尤度を γ のプロファイル尤度と呼ぶ。つまり、この尤度を最大にする $\hat{\gamma}$ が、求める最尤推定量であり、それは適当に用意した γ の数値列

$$\{\gamma_1 \leq \gamma_2 \leq \dots \leq \gamma_M < \min x_i\}$$

に対応したプロファイル尤度 $\ln L^{**}(\gamma_j, j=1, \dots, M)$ を計算して最大値を探す一次元数値探索法で簡単に求められる。また、この方法を利用すると、 γ の両側 $(1-\alpha)$ 水準の信頼区間 (Profile likelihood based confidence interval) が次式を満足する γ の区間

$$\left\{ \gamma: \ln L^{**}(\gamma) \geq \ln L^{**}(\hat{\gamma}) - \frac{1}{2} \chi^2_i(\alpha) \right\} \quad (6)$$

として容易に推定できる。ここに、 $\chi^2_i(\alpha)$ は自由度 1 の χ^2 分布の上側 100α パーセントイルである。

対数正規分布は一般に高値に裾を長く引く非対称な分布形状を示すが、正規分布のように対称性に近い分布を示すものも少なくない。このような場合には、正規分布との判別が不可能となり式(5)のプロファイル尤度関数が γ の単調減少関数となり $\hat{\gamma} \rightarrow -\infty$ となる。したがって、このような形状を示す場合には曝露日の推定は統計学的にはできないことに注意したい。正規分布の左側の裾が $-\infty$ への伸びていることを考えれば、この結果は容易に想像できるだろう。ところで、従来の方法では、対数正規分布とは言えないようなデータに対しても盲目的に「推定値」を計算してしまう危険性が高い。ここにも従来の方法の不適切性の別の側面が表れていると言えよう。

ところで、上記に提案した方法は観測値に連続値を仮定したものであるが、実際の集団食中毒の疫学調査で観測されるデータは何日に何例という区間頻度データであることが多い。したがって、厳密には観測日の中点(階級値)を $\{m_i\}$, $i=1, 2, \dots, K$, とすると区間 $[m_i - 1/2, m_i + 1/2)$ に f_i 例, $\sum f_i = n$, という区間頻度データに基づく多項

分布の尤度を計算する必要がある^{14,15)}。しかし、実際には問題となるほど大きな影響はなく、また、この計算には積分を含む厄介な計算が必要であるので、推定値のバイアスの分析を、次に示すようなシミュレーションで検討する方が簡便で実用的であろう。

1. まず度数分布の各階級 $[m_i - 1/2, m_i + 1/2)$ の f_i 例のそれぞれのデータに区間 $(-1/2, 1/2)$ の一様乱数 u_{ij} を加えて連続変数 x_{ij}

$$x_{ij} \leftarrow m_i + u_{ij}, \quad i=1, \dots, K; j=1, \dots, f_i$$

を生成し、 n 個のデータ $\{x_{ij}\}$ に基づく γ の最尤推定値と95%信頼区間を求める。もちろん、四捨五入するともとのデータに一致する。

2. この計算を100回繰り返す

ここで、繰り返しの回数を100としたのはバイアス推定値の誤差の大きさが推定値の $1/\sqrt{100}=1/10$ のオーダーに押えられるという意味で本シミュレーションの目的には十分と考えたからである。ただ、いたずらに観測時間単位の何倍もの区間をとるとその差は無視できなくなるので、可能な最小観測時間間隔を利用したい。

IV 腸管出血性大腸菌による集団食中毒への適用例

ここでは、長野県衛生公害研究所の和田正道主任研究員より教えていただいた事例報告の中から生データが掲載されていた4事例、平成8年に発生した腸管出血性大腸菌による集団食中毒の報告事例3件(O-157:H7が2件、O-118:H2が1件)と平成5年に発生したO-157:H7による同報告事例1件、のデータに本方法を適用する。また、従来の方法の中で自動的に計算できて比較的良好に利用されているSartwellの方法も適用する。なお、この推定量は式(2)で $\hat{\gamma}_s(p)$ と与えられるように、未知のパラメータ p を選択しなければならない。ここでは文献的によく利用されている p の値0.15を中心にその前後の2つの値(0.1と0.2)、計3種類の値を利用した結果を示す。

1. 岡山県邑久町の事例

まず、平成8年5月に岡山県邑久町の小学校で発生した学校給食が感染源とみられるO-157:H7による集団食中毒事件の発症日別度数分布に提案する方法を適用してみよう。この食中毒では、脳症で児童2人が死亡している。データは市

場¹⁶⁾の図1より引用したもので表1に示した。計算のため、4月30日正午を原点 $x=0$ としよう。例えば、5月24日正午は $x=24$ であり、6月1日は $x=32$ である。実際の計算では、 γ の数値列を

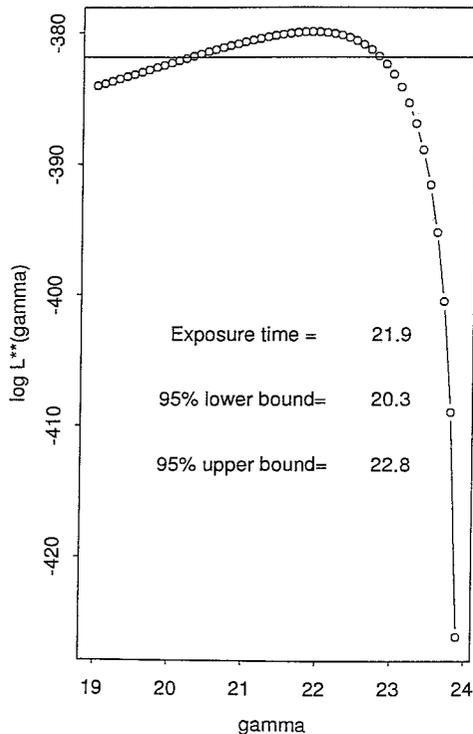
$$\gamma = \frac{j}{10}, \text{ for } j=190, 191, \dots, 239$$

として少数点以下一桁の精度（すべての事例に共通）でプロファイル尤度 $\ln L^{**}(\gamma)$ を計算した結果を図1に示した。曝露時点の最尤推定値はプロファイル尤度が最大となる γ の値で、 $\hat{\gamma}=21.9$ 、最大値は $\ln L^{**}(21.9) = -379.93$ であった。すなわち、5月22日の給食が最も疑われる。図1には x 軸に平行な線が描かれているが、これは

$$\gamma = \ln L^{**}(\hat{\gamma}) - \frac{1}{2} \chi_1^2(0.05) = -381.85$$

となる線である。式(6)より、この水平線と $\ln L^{**}(\gamma)$ との曲線との交点が95%信頼限界を与える。このようにして求めた95%信頼区間は(20.3, 22.8)であった。つまり、確率95%で曝露

Fig. 1 The profile likelihood $\ln L^{**}(\gamma)$ and the horizontal line indicated by $\gamma = \ln L^{**}(\hat{\gamma}) - 1/2\chi_1^2(0.05)$ for data from an outbreak of *Escherichia coli* O-157:H7 in Oku-cho, Okayama prefecture, Japan, 1996.



時点が5月20日から5月23日までの学校給食のいずれかであると推測できる。市場¹⁶⁾は学校給食の料理の調査から曝露日として5月22日と23日が最も疑われるとしているが、原因食品の特定は難航していると報告している。

さて、対数正規分布の適合度を視覚的に検討するために、

1. 発症日の相対度数分布の図上に推定された対数正規分布を重ね合わせた図を図2(a)、
2. 推定された潜伏期間の対数変換値、 $\log(x - \hat{\mu})$ 、の正規プロットを図2(b)、にそれぞれ示した。図2(a)の視覚的適合性と図2(b)の点の並びのほぼ直線性により対数正規分布への適合度はまずまずということになる。区間頻度データに基づいているので正規プロットの点が階段関数となっていることに注意したい。

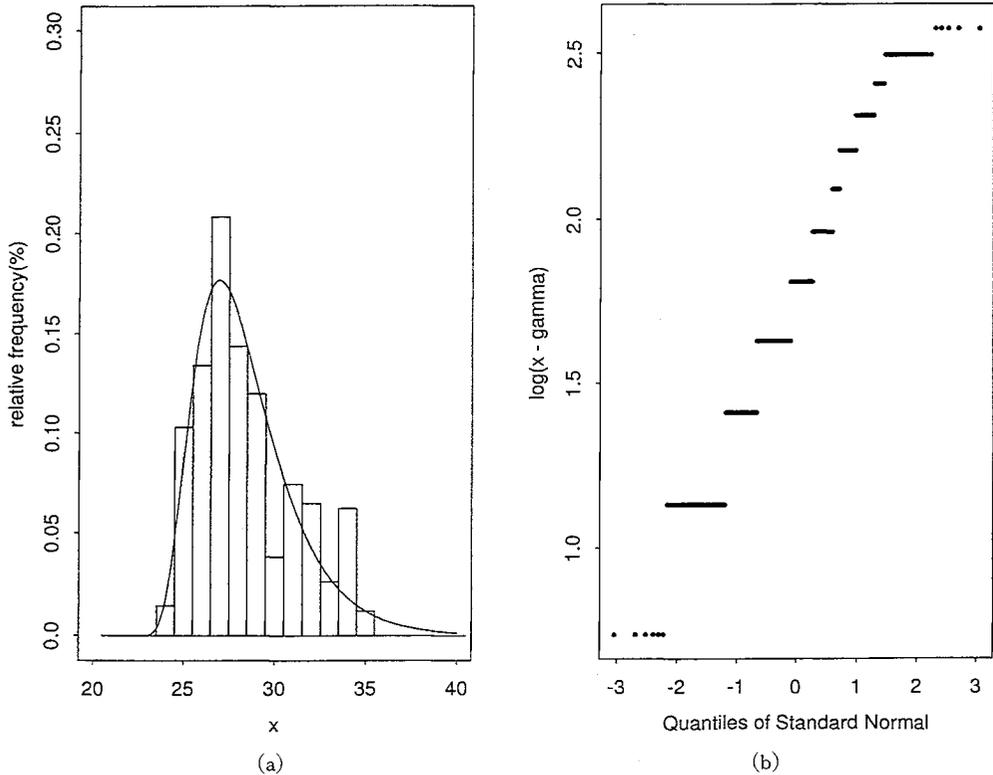
他のパラメータは $\hat{\mu}=1.79$ 、 $\hat{\sigma}=0.41$ と推定された。また、頻度データに基づく最尤推定値のもつバイアスを検討したシミュレーションの結果は、曝露日の推定値 $\hat{\gamma}$ の平均は $21.6 \pm 0.025 (S.E)$ となり、また、95%信頼区間の平均も $[19.8 \pm 0.018, 22.5 \pm 0.038]$ となった。頻度データに基づく推定値は0.3程度、高値方向に偏りが認められるが、結果の解釈には大きな違いはない。

ところで、Sartwellの方法を適用した結果は、 $\hat{\gamma}_S(0.1)=18.1$ 、 $\hat{\gamma}_S(0.15)=24.0$ 、 $\hat{\gamma}_S(0.2)=22.0$ 、とパラメータ p の選択によって大きく変化してしま

表1 平成8年5月岡山県邑久町の小学校で発生したO-157:H7による集団食中毒における発症日別度数分布(市場, 1996)

発症日	人数
5月24	6
25	43
26	56
27	87
28	60
29	50
30	16
31	31
6月1	27
2	11
3	26
4	5
計	418

Fig. 2 Two kinds of visual tools of for goodness-of-fit of the fitted lognormal distribution to data from an outbreak of *Escherichia coli* O-157:H7 in Oku-cho, Okayama prefecture, Japan, 1996: (a) Relative frequency distribution of onset dates and the fitted lognormal distribution, and (b) the normal probability plot of log-transformed estimated incubation periods, $\log(x-\hat{\gamma})$.



うのが観察される。

2. 岐阜市の事例

次に、平成8年6月に岐阜市の小学校で発生した学校給食が感染源とみられるO-157:H7による集団食中毒事件の初診日別医療機関受診者数に提案する方法を適用してみよう。データは河合ら¹⁷⁾の表1より引用したもので表2に示した。計算のため、5月31日正午を原点 $x=0$ としよう。例えば、6月1日正午は $x=1$ である。実際の計算では、 γ の数値列を

$$\gamma = \frac{j}{10}, \text{ for } j = -40, -39, \dots, 69$$

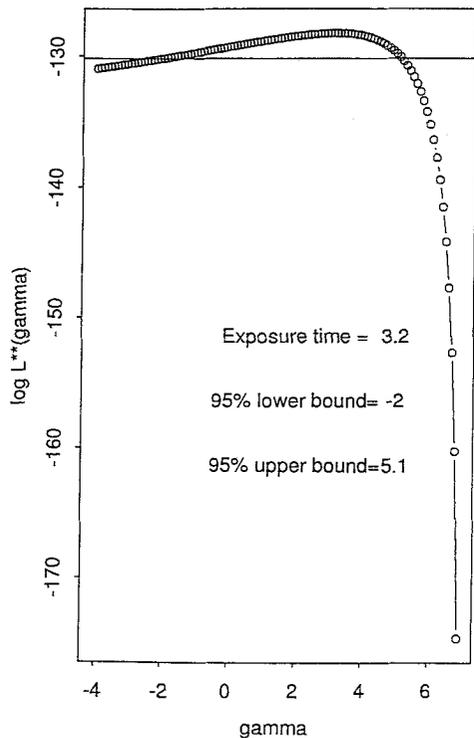
としてプロファイル尤度 $\ln L^{**}(\gamma)$ を計算した結果を図3に示した。曝露時点の最尤推定値は $\hat{\gamma}=3.2$ 、すなわち、6月3日の給食が最も疑われる。また、95%信頼区間は $(-2.0, 5.1)$ であった。つまり、確率95%で曝露時点が5月29日から

表2 平成8年6月岐阜市の小学校で発生したO-157;H7による集団食中毒における初診日別医療機関受診者数、(河合ら、1996)

初診日	受診者数
6月7(金)	2
8(土)	2
9(日)	3
10(月)	50
11(火)	55
12(水)	24
13(木)	29
14(金)	17
15(土)	7
16(日)	1
17(月)	8*
18(日)	2
計	200

* : 二次感染が疑われる受診者が多い

Fig. 3 The profile likelihood $\ln L^{**}(\gamma)$ and the horizontal line indicated by $\gamma = \ln L^{**}(\hat{\gamma}) - 1/2\hat{\sigma}^2(0.05)$ for data from an outbreak of *Escherichia coli* O-157:H7 in Gifu-shi, Gifu prefecture, Japan, 1996.



6月5日までの学校給食のいずれかであると推測できる。この信頼区間の幅は邑久町の場合に比較すると少々大きい。河合ら¹⁷⁾は、都立衛生研究所の検査で6月5日の給食（おかかサラダ）からO-157:H7が検出されたことから6月5日の曝露が疑わしいとしているが、野外実習のため5日の給食を食べていない5年生からも血便陽性者が1人検出されていることから、断定はできないだろう。

図4(b)に示した推定された潜伏期間の対数変換値の正規プロットでは、まずまずの直線性であるが、下部で上に凸な傾向を示す。図4(a)に示した初診日の相対度数分布と推定された対数正規分布の関係をみると9日から11日頃の間には適合が強いことからその原因がある程度推測できよう。つまり、9日日曜日に少なく、10日月曜日に多いという傾向が顕著である。これは、日曜日に受診できなかった有症者が月曜日に集中した可能性が十分考えられる。また、17日（月）は二次感

染が疑われる患者が多いとのことである。これらの要因は対数正規分布の適合度とその結果としての推定値に少なからず影響を与えたことは否定できない。特に、本報告例のように初診日のデータは曜日の影響を直接に浮けるので、調査では発症日のデータを収集すべきである。

他のパラメータは $\hat{\mu}=2.12$, $\hat{\sigma}=0.23$ であった。ところで、邑久町の事例で行ったのと同様のシミュレーションの結果は次の通りである：曝露日の推定値 $\hat{\gamma}$ の平均は 2.91 ± 0.05 (S.E) となり、また、95%信頼区間の平均も $[-2.54 \pm 0.10, 4.95 \pm 0.03]$ となった。結果の解釈には大きな違いがない。

次に、Sartwellの方法を適用した結果は、 $\hat{\gamma}_s(0.1)=9.5$, $\hat{\gamma}_s(0.15)=9.5$, $\hat{\gamma}_s(0.2)=9.0$, となった。今度はパラメータ p の選択によって大きく変化していないものの3つの推定値とも、初発患者が出現した7日より2日以上も後の時点を暴露時点と推定してしまっている。これは明らかに推定値として適切でないことを端的に表している事例であろう。

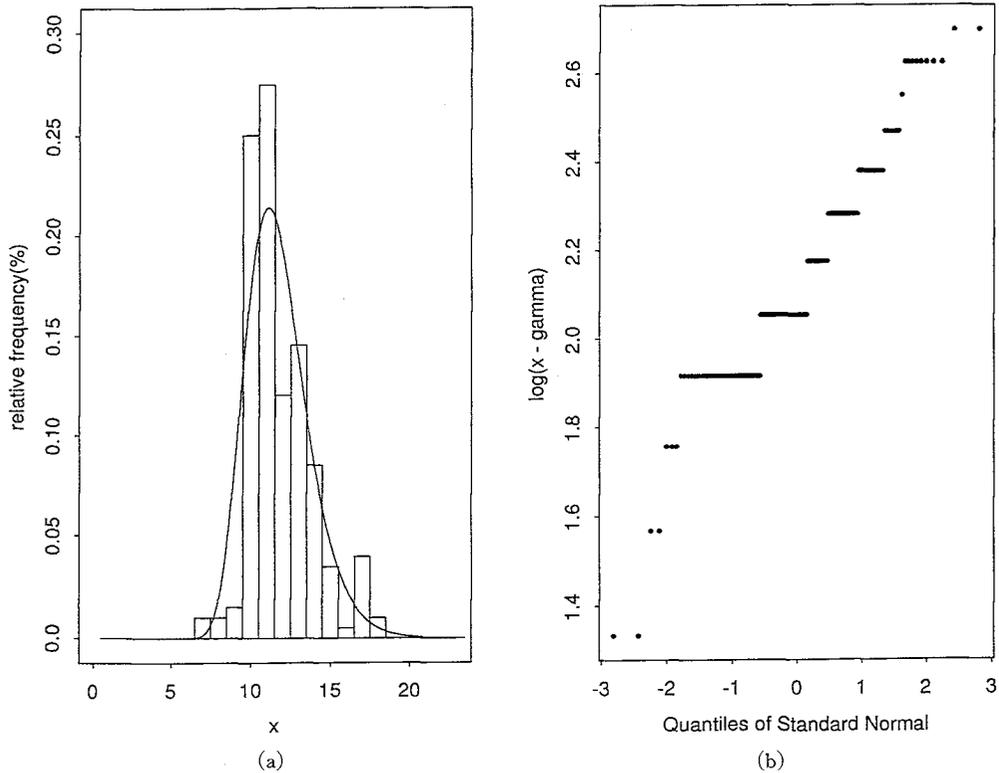
3. 石川県の事例

3番目は、平成8年7月に石川県の中学校で発生した学校給食がやはり感染源とみられる集団食中毒事件である。この原因菌はO-157:H7ではなくO-118:H2であった。使用するデータは初発症状の日別推移のデータで、蓮井ら¹⁸⁾の図2より引用したもので表3に示した。計算のため、6

表3 平成8年7月石川県の中学校で発生したO-118:H2による集団食中毒における発症日別度数分布（蓮井, 1997）

発症日	人数
7月10(水)	1
11(木)	2
12(金)	15
13(土)	22
14(日)	17
15(月)	10
16(火)	2
17(水)	5
21(日)	1
22(月)	1
計	76

Fig. 4 Two kinds of visual tools for goodness-of-fit of the fitted lognormal distribution to data from an outbreak of *Escherichia coli* O-157:H7 in Gifu-shi, Gifu prefecture, Japan, 1996: (a) Relative frequency distribution of dates of the first-visit to nearby hospitals and the fitted lognormal distribution, and (b) the normal probability plot of log-transformed estimated incubation periods, $\log(x-\hat{\gamma})$.



月30日正午を原点 $x=0$ としよう。例えば、7月1日正午は $x=1$ である。実際の計算では、 γ の数値列を

$$\gamma = \frac{j}{10}, \text{ for } j=20, 21, \dots, 99$$

としてプロファイル尤度 $\ln L^{**}(\gamma)$ を計算した結果を図5に示した。曝露時点の最尤推定値は $\hat{\gamma}=8.0$ 、すなわち、7月8日の給食が最も疑われる。また、95%信頼区間は (4.0, 9.3) であった。つまり、確率95%で曝露時点が7月4日から9日までの学校給食のいずれかであると推測できる。蓮井ら¹⁸⁾は「7月10日頃の給食と深い関連があったと推定せざるを得なかった」としているが、その推定根拠は明確でなく、感染源も特定できていない。

他のパラメータは $\hat{\mu}=1.70$ 、 $\hat{\sigma}=0.31$ であった。また、図6(a)(b)には発症日の相対度数分布の図

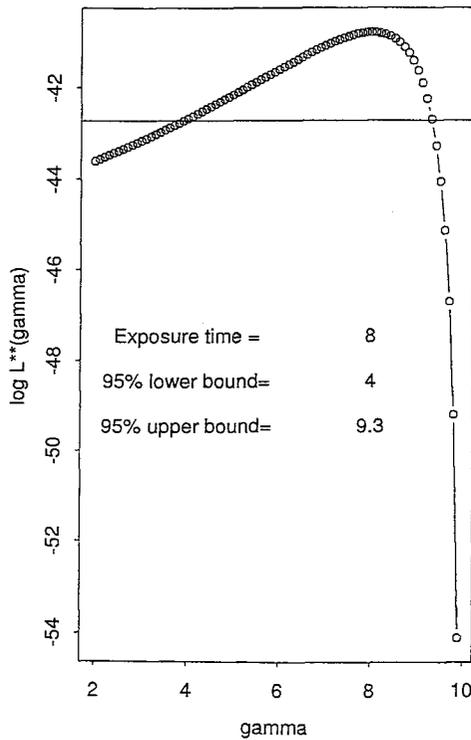
上に推定された対数正規分布を重ね合わせた図と推定された潜伏期間の対数変換値の正規プロットを示す。適合度はまずまずであろう。バイアス検討のためのシミュレーションの結果は次の通りである：曝露日の推定値 $\hat{\gamma}$ の平均は 7.80 ± 0.08 (S.E) となり、また、95%信頼区間の平均も $[3.46 \pm 0.15, 9.16 \pm 0.05]$ となった。結果の解釈には大きな違いがない。

次に、Sartwellの方法を適用した結果は、 $\hat{\gamma}_s(0.1)=11.5$ 、 $\hat{\gamma}_s(0.15)=11.0$ 、 $\hat{\gamma}_s(0.2)=11.0$ となった。パラメータの選択によって大きく変化していないものの、今度も、初発患者が出現した10日より後の時点を暴露時点と推定している。

4. 平成5年東京都の事例

最後は、平成5年8月に東京都杉並区の区立保育園で発生したO-157:H7による集団食中毒事件である。発症日別推移の度数は山村ら¹⁹⁾の図1

Fig. 5 The profile likelihood $\ln L^{**}(\gamma)$ and the horizontal line indicated by $\gamma = \ln L^{**}(\hat{\gamma}) - 1/2\chi_1^2(0.05)$ for data from an outbreak of *Escherichia coli* O-118:H2 in Ishikawa prefecture, Japan, 1996.



より引用したもので表4に示した。本件は上記3事例とは異なり、プール水を介した感染が疑われた事例である。

計算のため、7月31日正午を原点 $x=0$ としよう。実際の計算では、 γ の数値列を

$$\gamma = \frac{j}{10}, \text{ for } j=200, 201, \dots, 279$$

としてプロファイル尤度 $\ln L^{**}(\gamma)$ を計算した結果を図7に示した。曝露時点の最尤推定値は $\hat{\gamma}=26.8$ ，すなわち、8月27日に曝露した可能性が最も疑われる。また、95%信頼区間は (22.8, 27.7) であった。つまり、確率95%で曝露時点が8月23日から28日までのいずれかであると推測できる。山村ら¹⁹⁾は発症のピーク、O-157の潜伏期間から経験的に判断して感染は8月23-25日頃と推測しているが、その根拠は極めて乏しい。

他のパラメータは $\hat{\mu}=1.29$ ， $\hat{\sigma}=0.51$ であった。

図8(a)(b)には、発症日の相対度数分布の図上に推定された対数正規分布の重ね合わせた図と推定

表4 平成5年8月東京都杉並区立保育園で発生したO-157:H7による集団食中毒における発症日別度数分布 (山村, 1996)

発症日	人数
8月28	2
29	4
30	11
31	7
9月1	1
2	1
3	1
5	1
6	2
計	30

された潜伏期間の対数変換値の正規プロットを示す。発症者数が30人と少ないのと離散分布であるため対数正規分布への適合度を視覚的に判断するのは容易ではないが、極めて不適合と言えるものではないことが観察される。ここでも、バイアス検討のためのシミュレーションの結果は次の通りである：曝露日の推定値 $\hat{\gamma}$ の平均は 26.6 ± 0.09 (S.E) となり、また、95%信頼区間の平均も $[22.1 \pm 0.29, 27.5 \pm 0.05]$ となった。やはり、結果の解釈には大きな違いがない。Sartwellの推定値は、 $\hat{\gamma}_S(0.1)=28.7$ ， $\hat{\gamma}_S(0.15)=28.4$ ， $\hat{\gamma}_S(0.1)=29.4$ ，となった。ここでも3種類の推定値すべてが、初発患者が出現した28日より後の時点を曝露時点と推定している。

5. 4事例に関する曝露時点と潜伏期間

まず、表5にこれら4つの集団食中毒事例における曝露時点の推定値をまとめた。区間データが最尤推定値に与えるバイアスは4つの事例とも小さいことが観察される。この4事例の推定値からでも明確に結論できることは、初発患者が出現した後の時点を曝露時点と推定してしまうことの多い「Sartwellの方法は利用すべきでない」ということである。

次に、表6に4事例における対数正規分布の曝露時点 γ 以外の2つのパラメータ (μ, α) の推定値とそれから計算される潜伏期間の100%点 T_α

$$T_\alpha = \exp(\hat{\mu} + z_\alpha \hat{\sigma})$$

を示した。ここに z_α は平均0分散1の標準正規分布の100%点である。岡山県の事例と石川県

Fig. 6 Two kinds of visual tools for goodness-of-fit of the fitted lognormal distribution to data from an outbreak of *Escherichia coli* O-118:H2 in Ishikawa prefecture, Japan, 1996: (a) Relative frequency distribution of onset dates and the fitted lognormal distribution, and (b) the normal probability plot of log-transformed estimated incubation periods, $\log(x - \hat{\gamma})$.

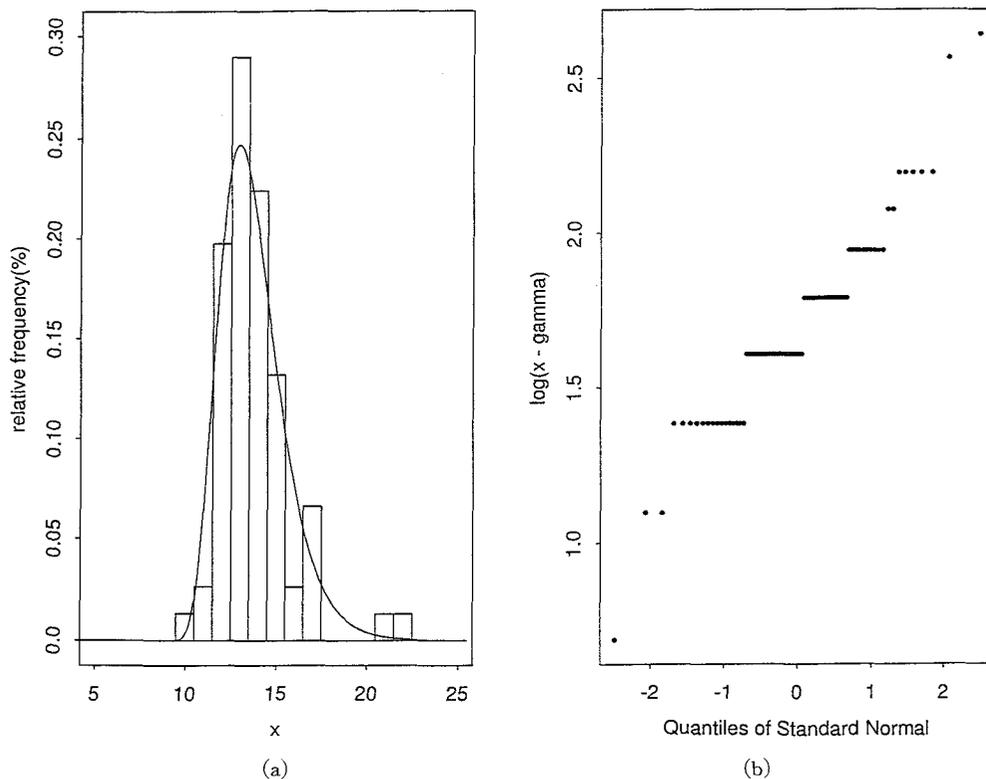


表5 4事例における曝露時点の推定値のまとめ

事例	最尤推定法		シミュレーション ^a		Sartwellの推定値 $\hat{\gamma}(\rho)$		
	$\hat{\gamma}$	95%信頼区間	$\hat{\gamma}$	95%信頼区間	p=0.1	p=0.15	p=0.2
岡山県	21.9	(20.3, 22.8)	21.6	(19.8, 22.5)	18.1	24.0	22.0
岐阜市	3.2	(-2.0, 5.1)	2.9	(-2.54, 4.95)	9.5 ^b	9.5 ^b	9.0 ^b
石川県	8.0	(4.0, 9.3)	7.8	(3.5, 9.2)	11.5 ^b	11.0 ^b	11.0 ^b
東京都	26.8	(22.8, 27.7)	26.6	(22.1, 27.5)	28.7 ^b	28.4 ^b	29.8 ^b

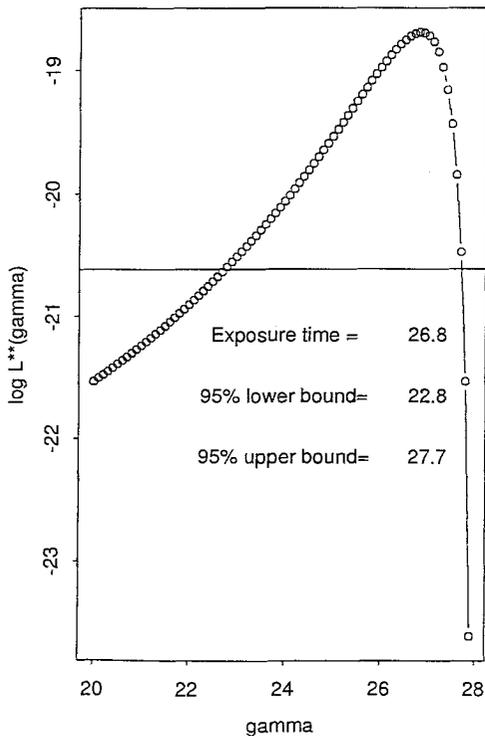
^a: 推定値の標準誤差は省略した。

^b: これらの推定値は初発患者が出現した日より後の時点である。

の事例では原因菌がそれぞれ O-157:H7, O-118:H8 と異なるが、推定値が大変類似している。岐阜市の事例は初診日のデータを利用しているため、日曜日に発症した患者が月曜日にカウントされている可能性が高いことを反映してメディアン潜伏期間が2, 3日長めと推定されている。また東京都の事例は30人と患者数が少ないことと、他

の事例と違って学校給食ではなくプール水を介した感染が疑われた事例であることから他の3つの事例と対等に比較はできないように思われる。これらの結果から潜伏期間が対数正規分布に従うという仮定が正しければメディアン潜伏期間は「5-6日」前後と推測されることになる。

Fig. 7 The profile likelihood $\ln L^{**}(\gamma)$ and the horizontal line indicated by $\gamma = \ln L^{**}(\hat{\gamma}) - 1/2\chi_1^2(0.05)$ for data from an outbreak of *Escherichia coli* O-157: H7 in Sugina-mi-ku, Tokyo, Japan, 1993.



V 考 察

一般にある統計モデルが実際の現象をよく記述できるかどうかを評価する場合

1. モデル（本論文では「潜伏期間が対数正規分布に従う」とする仮定）の妥当性
2. モデルに含まれるパラメータの推定法の良さ

の2つの要素に分けて議論する必要がある。後者は明らかに統計学の問題であるのに対し前者は統計学だけでは解決できないその分野固有の問題であり、現象への十分な理解と経験が不可欠である。

本論文の主要な目的は、後者であり潜伏期間に対数正規分布を仮定した場合の曝露時点の推定方法の良さを検討するとともに公衆衛生分野ではあまり知られていない最尤推定法を提案することにある。したがって、まず、これまで提案されている推定法を統計学的に評価した結果、そのすべて

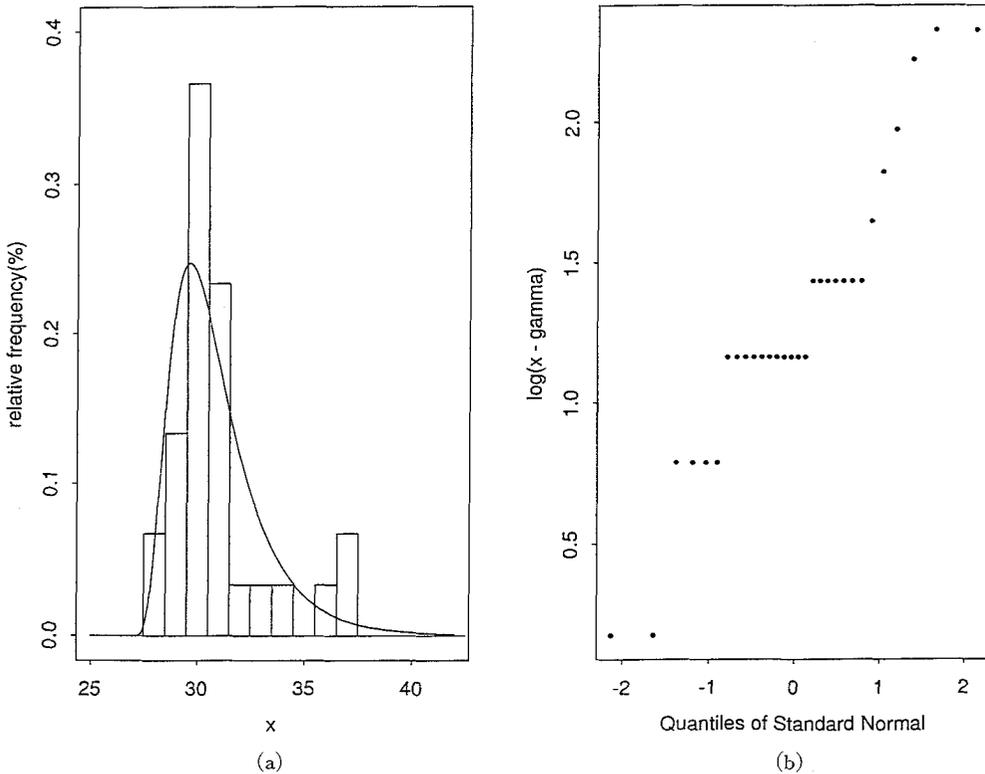
表6 4事例における対数正規分布のパラメータ (μ, σ) の推定値と潜伏期間の推定値

事例	患者数	対数正規分布		潜伏期間の推定値		
		$\hat{\mu}$	$\hat{\sigma}$	5%点	メデアン	95%点
岡山県	418	1.79	0.41	3.06	5.99	11.73
岐阜市	200	2.12	0.23	5.71	8.33	12.15
石川県	76	1.70	0.31	3.29	5.49	9.10
東京都	30	1.29	0.51	1.57	3.63	8.38

が統計理論上不適切か、あるいは、明らかに推定精度の悪いものであり、実用に耐えるものではないことを指摘した。また、自動的に計算可能な Sartwell の方法については実例でもその悪さを実証した。次に、これらに代わって、漸近的に最適な推定量が得られる最尤推定量の簡単な計算方法を提案し、プロファイル尤度に基づく信頼区間の計算方法も新しく提案した。本論文で示した最尤法の簡単な計算法は統計学に関連する文献でもほとんど知られていない^{10~14)}。また、これらの計算方法の統計学的な妥当性はすでにシミュレーション等ですでに確かめられている¹⁵⁾。本論文ではさらに、実際の調査データは表1のような区間頻度データであることが多いので、シミュレーションによる推定値のバイアスの簡単な検討方法も示した。ただ、そこで使用した一様乱数は対数正規分布の形状をなめらかに表現できないという意味では必ずしも適切ではない。隣接の階級度数を考慮にいたした矩形乱数などを利用したほうがより適切な検討が可能となるだろう。

本論文で提案した最尤法が他の3つの方法に比べると最も良い方法であることは理論的に明らかであるが、その方法の実際問題への適用にあたっては、本論文の主要目的ではないものの、モデルの妥当性、つまり、「腸管出血性大腸菌 O-157: H7 に代表される原因菌に一齐に曝露することによって発生する集団食中毒の潜伏期間が対数正規分布に従うか？」というより本質的な仮定の妥当性のある程度の検証は避けて通れない。そこで、本論文ではその検証の方法として「モデルのデータへの適合度」を視覚的に評価するために、(1)発症日の相対度数分布に推定された対数正規分布を重ね合わせた図、(2)推定された対数変換後の潜伏

Fig. 8 Two kinds of visual tools for goodness-of-fit of the fitted lognormal distribution to data from an outbreak of *Escherichia coli* O-157:H7 in Sugunami-ku, Tokyo, Japan, 1993: (a) Relative frequency distribution of onset dates and the fitted lognormal distribution, and (b) the normal probability plot of log-transformed estimated incubation periods, $\log(x-\hat{\gamma})$.



期間の正規プロット，の2種類の方法を利用した。本論文で解析した4件の集団食中毒事例では対数正規分布の適合度は岐阜市の事例を除けばまずまずであったと評価でき，対数正規分布が感染症の流行事例の潜伏期間の分布をよく記述できたという過去の経験^{1~3,6~8,20)}がここで通用したと評価できるかも知れない。

しかし，その生物学的根拠は依然希薄であるため推定された結果の解釈上の次のような問題は依然解消していない。潜伏期間の真の分布が対数正規分布に従えばデータも対数正規分布によく適合するが，その逆は必ずしも成立しない。つまり，モデルのデータへの適合度の良さは真の潜伏期間の分布が対数正規分布に従うための必要条件ではあるが十分条件ではないということである。したがって，対数正規分布の仮定が正しければ漸近的に最良の推定値を与える本方法も，対数正規分布

の仮定が正しくない場合には，推定値の良さは保証できないのである。腸管出血性大腸菌 O-157:H7 による食中毒の潜伏期間のより生物学的に興味のある統計モデルを検討するためには，まず，適当な動物実験を行って正確な潜伏期間の分布型に関するデータを取る必要があるように思われる。

他方で，平成8年6月の岐阜市の事例のように発症日ではなく医療機関への初診日のデータを集計してしまうと日曜日には少なく，月曜日には多いといった本来感染とは無関係な「受診しやすい曜日」の影響を受けたデータとなってしまう。また，明らかに二次感染であると判定できる患者も除かなければならない。つまり，潜伏期間が推定できるデータを調査しないと，プロットした患者発生時間分布は一斉曝露による正しい流行状況を反映しなくなり，その結果，対数正規分布の仮定

は妥当であっても、分布型は対数正規分布に適合せず、真値からかけ離れた推定値が得られることにつながることに注意したい。もっとも、発症時間のデータを慎重に調査しても非現実的な推定値が得られた場合は対数正規分布というモデルの妥当性が強く疑われるのでその結果は無視すべきである。

一般に、集団食中毒の感染源の調査には、食品の細菌検査、ケース・コントロールを基本とした喫食調査などが平行して実施される。しかし、発生してからの後ろ向き調査であるだけに、曝露時点と感染源の特定に必要な決め手となるデータが時間が経過するにしたがって得にくく、困難であることが多い。本論文で紹介した集団食中毒の4事例でも、明快な論理で感染日と感染源を特定できているものはない。感染してから数時間後に食中毒症状が表れる潜伏期間の短い原因菌の場合は別であるが、腸管出血性大腸菌 O-157:H7 に代表される比較的潜伏期間の長い原因菌に基づく集団食中毒では、本アプローチによって得られた曝露時点の推定値とその信頼区間は、これらの調査からは得られない有用な情報を提供できる可能性があるものと思われる。また、表6に示した潜伏期間とそれを推定するためのパラメータ ($\hat{\mu}$, $\hat{\sigma}$) に関するデータは今後の腸管出血性大腸菌を原因とする食中毒の潜伏期間を推測する上で貴重である。本方法で他の事例を解析していくことによりその潜伏期間に関する情報が蓄積されていくことになる。そうすれば、この事前情報を利用したベイズの推論 (Bayesian inference) が可能となりより妥当な曝露時点の推測が可能となるだろう。

最後に、本論文で提案した方法は、潜伏期間の分布に対数正規分布以外の、例えば、ワイブル分布、ガンマ分布、などの多くの確率分布に容易に拡張することができる。興味があるのは最適モデル選択の問題であるが、今後の課題としたい。

本論文で提案した方法の計算プログラムは統計ソフト S-PLUS を利用した。このプログラムを試してみたい読者には、学術的目的使用の範囲で提供する用意があるので、希望者は著者まで。なお、email は「tango@iph.go.jp」。

本研究の一部は平成9年度文部省科学研究費補助金基盤研究 C (2) 課題番号09680319で行った。

(受付 '97. 4. 15)
採用 '97. 11. 20)

文 献

- 1) Sartwell PE. The incubation period and the dynamics of infectious disease. *Am J Epidemiol* 1966; 83: 204-216.
- 2) 平山 雄. 疫学, 績文堂, 1958; 82-83.
- 3) 堀内一弥, 中井清三, 上嶋 勲, 杉山 博. 曝露時点の推定に関する理論疫学的研究 (その 1). *日本公衛誌*, 1956; 3: 184-186.
- 4) Meynell GG. Interpretation of distribution of individual response times in microbial infections. *Nature* 1963; 198: 970.
- 5) 金光正次, 他. 疫学とその応用, 第3版, 南山堂, 1970; 87-89.
- 6) 中村恭政. 食中毒・経口伝染病流行時におけるいわゆる感染曝露時点の推定方式に関する比較研究. *日本公衛誌*, 1964; 11: 85-114.
- 7) 山本俊一. 疫学総論, 文光堂, 1970; 319-325.
- 8) 坂本洲弘. 疫学と疫学モデル. 金芳堂, 1985; 297-298.
- 9) 柳川 洋編. 疫学マニュアル, 南山堂, 1988; 101.
- 10) Aitchison J and Brown JAC. *The Lognormal Distribution*, London, Cambridge University Press. 1957.
- 11) Johnson NL and Kotz S. *Continuous Univariate Distributions-1*. London, John Wiley & Sons. 1970.
- 12) Cohen AC. Three-parameter estimation. in "Lognormal Distribution-Theory and Applications", eds. by Crow EL and Shimizu K, New York, Marcel Dekker. 1988; 113-137.
- 13) Cohen AC and Whitten BJ. Estimation in the three-parameter lognormal distribution. *J Ame Statist Ass* 1980; 75: 399-404.
- 14) Giesbrecht F and Kempthorne O. Maximum likelihood estimation in the three-parameter lognormal distribution. *J R Statist Soc B* 1976; 28: 257-264.
- 15) Tango T. On the simplest maximum likelihood estimator in the three-parameter lognormal distribution: motivated by school-lunch-related outbreaks of *Escherichia Coli O-157: H7*. *日本計量生物学会・応用統計学会1997年度合同年次大会講演集*. 1997; 101-106.
- 16) 市場洋三 (1996), 岡山県における腸管出血性大腸菌 O-157:H7 による集団食中毒の概要—国立岡山病院小児科はいかに対応したか—, *日本医事新報*, 1996; 3785: 26-30.
- 17) 河合直樹, 広瀬 洋, 岩砂和雄, 青木 靖, 山崎喜久, 臼井曜子, 松浦省三. 岐阜市における腸管出

- 血性大腸菌 O-157による集団食中毒について—概要と緊急アンケート結果—, 日本医事新報, 1996; 3771: 27-30.
- 18) 蓮井正亮, 加登康洋, 杉宮貞栄, 水腰久美子, 西正美. 腸管出血性大腸菌 O-118:H2による集団下痢症(速報)—石川県 R 中学校—, 日本医事新報, 1996; 3788: 27-29.
- 19) 山村朋子, 織田壽昭, 森谷憲治, 前田 正, 金児克忠, 鈴木末男, 小川幹雄, 森田師郎, 根本信司. 区立保育園で発生した腸管出血性大腸菌(O-157:H7)による集団下痢症について, 食品衛生研究, 1996; 42: 89-95.
- 20) Sartwell, PE. The Distribution of incubation periods of infectious diseases. *Ame J Hygiene* 1950; 51: 310-318.

MAXIMUM LIKELIHOOD ESTIMATION OF DATE OF INFECTION IN AN OUTBREAK OF DIARRHEA DUE TO CONTAMINATED FOODS ASSUMING LOGNORMAL DISTRIBUTION FOR THE INCUBATION PERIOD

Toshiro TANGO*

Key words: Confidence interval, *Escherichia coli* O-157: H7, Maximum likelihood estimation, Profile likelihood

This paper proposes a new method for estimating the date of infection in an outbreak of diarrhea caused by eating foods contaminated by some sort of foodborne pathogens such as *Escherichia coli* O-157: H7. Simple methods for calculating maximum likelihood estimator and confidence interval are proposed assuming (1) a common date of infection and (2) the incubation period has a log-normal distribution. Further, this paper shows that all three methods proposed so far for estimating the date of infection under the same assumptions are theoretically inappropriate or computationally imprecise. The proposed methods are illustrated with data from three outbreaks of *Escherichia coli* O-157: H7 and outbreak due to *Escherichia coli* O-118: H2 in Japan.

* Division of Theoretical Epidemiology, Department of Epidemiology, National Institute of Public Health