

19990090

健康危機関連統計の高度処理に関する研究

(研究課題番号：H10-統計-005)

平成11年度厚生科学研究費補助金

(統計情報高度利用総合研究事業)

研究報告書

平成12年4月

研究代表者 金藤浩司

(統計数理研究所・統計科学情報センター)

厚生科学研究費補助金（統計情報高度利用総合 研究事業）

総括研究報告書
健康危機関連統計の高度処理に関する研究
主任研究者 金藤浩司 統計数理研究所

研究要旨 健康危機関連情報、特に腸管出血性大腸菌 0-157 の発生データに焦点を絞り、患者の発生予測と散在的集団発生に関する統計理論を構築するとともに、データの視覚化に関する検討を行い地理情報システムの可視化ツールとしての有効性を検証した。

分担研究者 越智義道
大分大学助教授
岩瀬晃盛
広島大学教授
馬場康維
統計数理研究所教授

A 研究目的

従来の統計情報の公開方法のみならず、緊急時において国民・地方自治体等から、迅速でかつ質の高い統計情報提供の要望がある。本研究では、腸管出血性大腸菌 0-157（以下では 0-157 と省略する。）を原因とする食中毒の発生に関して厚生省が保有する統計情報を処理し、国民・地方自治体等に質の高い統計情報を伝達する基礎となる統計解析システム構築に向けた基礎的研究を目的とする。

本研究から得られる成果をより発展させ、健康危機関連情報の時間・空間的解析等を行うことにより、疾患の経年変化や地域特性を明らかにすることが出来る。これにより予測を含め新たな事象が発生した場合への迅速な対応が可能となり、質の高い統計情報が提供可能となる。

B 研究方法

本研究は、大きく分けると 0-157 の発生情報に関する統計解析手法の構築と発生状況の可視化に関する検討という 2 つの基礎的研究から成る。

統計解析手法としては、0-157 を原因とす

る食中毒の発生パターンは、逆ガウス型分布等の寿命分布を用いて記述可能である場合が多い。そこで、統計理論の構築は、この仮定に基づいて進めた。逆ガウス型分布は、正のドリフトを持つウィーナー確率過程のある値への初期到達時間分布として導出される。このような分布自身が有するメカニズムが、本研究では重要な意味を持っている。

可視化に関する検討としては、前年に引き続き動的グラフィカル・データ解析システム（開発環境 Unix）を用いた動的表示の可能性について検討を加えるとともに、新たに地理情報システム（GIS）を用いたデータの表示能力と処理能力の検討を行った。今回の GIS システムとしては解析システムとの連携の観点から ESRI 社の ArcView を採用した。さらに、分析結果の公開の観点から、インターネットを利用した情報提供の可能性についても検討を行った。

C 研究結果

患者の発生パターン予測問題についての検討と、それに伴う統計解析手法の構築（2 母数逆ガウス型分布に従う 2 つの母集団における母平均、母変動係数に関する検定、微分方程式を利用した発生パターン曲線の算出）を試み、理論の構築とともに、データ解析プログラム（先に述べた手法以外に、2 母数逆ガウス型確率紙、3 母数逆ガウス型分布の未知

母数の最尤推定を含む。)として Mathsoft 社の S Plus+StatServer 上での実装の検討を行った。

表示については基本的には静的なグラフィカル表示に依存するものの、背景となる関連情報やデータ管理に有効と言われている GIS システムを採用して、0-157 関連情報の提示可能性について検討した。GIS ではデータの管理について表計算プログラムやデータヘースシステムとの連携が容易であり、関連情報の連結表示、地図上でのデータの選択表示、あるいは逆に地図上のデータによるデータ選択と詳細情報の確認に有効に機能することを確認してきた。また、外部プログラムとの連結に関する拡張機能を利用し、データ解析プログラム S-Plus との連携も可能である。さらに、これらの発生状況情報をインターネットを介して公開するには、ESRI 社が無料で提供している地図データ閲覧ソフト ArcExplorer を利用すれば、上記 GIS システムで生成した地理情報ファイルを提供することによって技術的に可能であることを確認した。

D 結論

統計解析手法の基礎的研究としては、現状においてデータ収集後に、手元にある情報から如何にして、対策の基礎資料となる情報が抽出可能か検討した。

0-157 データを取り扱う場合、以下の2つの問題点がある。

- 1) 発症日か同定できない個体が多く存在する。
- 2) 集団としてのデータが確定されるまでに時間がかかる。

これらの問題点を解決するためには、さらなる統計手法の構築が必要である。そのためにも、基礎情報として、より詳細なデータか、個人のプライバシーが守られた形で、迅速に利用できる環境が必要である。また、本研究をさらに進めることによって、厚生省が構築している FoodNet 等との相互補完的な関係が可能である。

一方、計算機環境の急速な発展により地理情報システムの利用も動的なグラフィカル・データ解析環境もようやく汎用的な手段として社会的認知を得て来たところであり、これからこれらの手法の真価を発揮できる場が多くなるものと考えられる。その際にデータの迅速で正確な確保は重要な問題となってくる。

E 研究発表

1 論文発表

R P Galapate, A U Baes, K Ito, K Iwase and M Okada (1999)、Trihalomethane formation potential prediction using some chemical functional groups and bulk parameters, Water Res Vol 33, No 11, pp 2555-2560

2 学会発表

金藤 浩司、岩瀬 晃盛、Estimation for a scale parameter with known coefficient of variation、第67回日本統計学会講演報告集、pp 75-76 1999 7 29

越智義道、金藤浩司、健康危機管理関連情報の視覚化とその分析、第67回日本統計学会講演報告集、pp 277-278、1999 7 30

横井浩文、小畑経史、越智義道、二つの二項比率の比較問題における条件付検定と条件なし検定の比較、第21回大分統計談話会 2000 2 18

馬場康維、地理情報を用いた統計データの利用、第67回日本統計学会講演報告集、pp 273-274、1999 7 30

Y Baba、The present and future of statistical education、International Statistical Forum、Korea、1999 9 17

中村好宏、馬場康維、大隅昇、損失関数最小化による射影と変数の分類、ISM Symposium、Recent Advances In Statistical Research and Data Analysis、Tokyo、2000 3 22

厚生科学研究費補助金（統計情報高度利用総合 研究事業）

分担研究報告書

健康危機関連統計情報の視覚化システムに関する基礎的研究

分担研究者 越智義道 大分大学

研究要旨 地理情報システムを用いることで厚生省が公開している腸管出血性大腸菌 O-157 を原因とする食中毒患者の発生情報を、詳細に把握できること等が明らかになった。

A 研究目的

本研究は健康危機関連統計情報について、その情報を分析し適切に状況を把握することを支援するためのシステムの構築に関する検討とその提案ならびにその問題点を把握することを目的とする。

厚生省では健康情報の一つとして O-157 などの食中毒に関する情報をホームページ上で公開していた。ここではこのようなデータが入手可能である場合にどのような形態でデータを把握し、それをどのような形で分析し情報の把握を行うことか可能であるかについて検討を加える。特にデータの視覚的表示の可能性について検討する。

B 研究方法

O-157 他食中毒等のデータでは、それら疾患の同定と発生経緯・原因の確認、さらにその予防などが問題となる。したがって分析の観点をとどこに置かかによって、そのデータとしての扱い方が変わってくるか、その時間情報、地理情報は非常に重要な情報といえる。

統計的なデータ解析に関する検討は、その時間情報に関する分析を主眼に置いたものであったか、それら解析対象となる対象者の選定やあるいは発生状況の確認のために、まず探索的な分析が必要とされる。探索的な分析では、データを視覚化表現し、解析者の視覚的な情報処理能力を利用することが有効であることか知られている。また本研究の対象となるような地理的な情報を含むデータの分析では、地理的特性やその地理的関連性を把握する上でその情報を視覚的に表現することが不可欠な要素となる。

ここではこのような地理的情報を含んだデータをグラフィカル表示することにより、経時変化・地域性などの特性を考慮に入れてこうした疾患の発生状況を分析し、その情報を提供することを目的とするシステムの構想について議論した。

前年に引き続き動的グラフィカル・データ解析システム（開発環境 Unix）を用いた動的表示の可能性について検討を加えるとともに地理情報システム（GIS）を用いたデータの表示能力と処理能力の検討を行った。今回の GIS システムとしては解析システムとの連携の点から ESRI 社の ArcView を採用した。さらに、分析結果の公開の観点から、インターネットを利用した情報提供の可能性についても検討を行った。

C 研究結果

表示については基本的には静的なグラフィカル表示に依存するものの、背景となる関連情報やデータ管理に有効といわれている GIS システムを採用して、O-157 関連情報の提示可能性について検討した。

GIS ではデータの管理について表計算プログラムやデータベースシステムとの連携が容易であり、関連情報の連結表示、地図上でのデータの選択表示、あるいは逆に地図上のデータによるデータ選択と詳細情報の確認に有効に機能することが確認できた。また、外部プログラムとの連結に関する拡張機能を利用し、データ解析プログラム S-Plus との連携も可能である。

さらに、これらの発生状況情報をインターネットを介して公開する準備として、ESRI 社が無料で提供している地図データ閲覧ソフト ArcExplorer を利用すれば、上記 GIS システ

ムて生成した地理情報ファイルを提供することによって技術的に可能であることを確認した。

D 結論

特に位置情報を含むデータについては、近年の地理情報システムの急速な普及により、これらの地理情報表示機能に特化した特性を有効に活用してデータの解析を進めることが必要である。また、地理情報システムではその同定機能や、選択機能の充実に目向けられがちであるか、空間的な情報把握として運用を図り、さらに関連データと有機的に連携してデータ解析に反映させることが重要である。

E 研究発表

1 口頭発表

越智義道、金藤浩司、健康危機管理関連情報の視覚化とその分析、第67回日本統計学会講演報告集、pp 277-278、1999 7 30
横井浩文、小畑経史、越智義道、二つの二項比率の比較問題における条件付検定と条件なし検定の比較、第21回大分統計談話会 2000 2 18

分担研究報告書

健康危機関連統計情報の統計処理に関する基礎的研究

分担研究者 岩瀬晃盛 広島大学

研究要旨 健康危機関連情報、特に腸管出血性大腸菌 0-157 の発生データに焦点を絞り、患者の発生予測と散在的集団発生に関する統計理論の構築を行った。

A 研究目的

一般に、0-157 を原因菌とする集団食中毒の発生は、ほぼ同時期に原因食材を摂取した患者か、平均して数日後に食中毒の症状を訴えることによって確認される。ここで公衆衛生上問題となる点は、0-157 を原因とする食中毒の確認までの期間に個体差があり、他の原因菌を原因とする食中毒に比べ食中毒の発生の確定までの時間が長い点である。このことが原因食材の推定や確定を難しくしている。これは、相当数の児童生徒が同じ食材を摂取する学校給食を原因とした場合においても、最終的な原因食材の確定を困難にしている要因である。

以上のような状況において、最終的な医学・細菌学・疫学的見地からの原因食材等の確定は必要であるが、人的・物的資源の問題により、各地で発生している集団食中毒への迅速な対応には至っていないと思われる。そこで現状の行政システムで得られるデータを効果的に利用するため、探索的・仮説検証的な手順として、統計解析手法が有効であり、同時に統計情報資源の有効活用の意味からも重要である。

そこで、患者の発生予測と散在的集団発生に関する統計理論の構築を目的とした。

B 研究方法

逆ガウス型分布は、正のドリフトを持つウィーナー確率過程のある値への初期到達時間分布として導出される。このような分布自身が有するメカニズムが、本研究では重要な意味を持っている。

0-157 データにおける患者の発症パターンのメカニズムを次のように考える。独立な個体が数多く居るところに共通の要因で病原菌

が摂取され感染したとしたときに、個体の代謝能力やその他生活環境等の相違により発症までに要する時間に個体差が生じる。

ここでは、感染して直後に発症する人をも確率が極めて低いにせよ想定している。また、決定論的に考えて感染してからの時間に比例して病原菌が増殖して、ある閾値に達したら発症すると考える。また、完全に同しいわばクローン人間の集団であると考えると全く同じ生活環境であるとすれば、感染者は全く同じ時刻に発症する。この時刻が2母数逆ガウス型分布で言えば平均値に相当する。

実際には、個体は同じでなく、生活環境も異なるから発症までの時間に変動が生じ、発症時間分布が逆ガウス型分布に従うと考えるのが妥当である。

C 研究結果

原因食材がある程度、確定されれば当然それを食べた集団の人数もある程度把握可能であるが、0-157 を原因物質とする集団食中毒の場合、発症までの時間が平均すると数日必要であり、その結果原因食材に関する確定が難しく、それに伴ってどの程度の人が原因物質を含む原因食材を摂取しているかは、到底早期に確定出来ない。

このような状況では、ある時点までの発症者の初症状の発症時刻に関するデータのみに基づいて、それ以降の発生状況を予測し、行政としての対策を施す必要がある。

そこで、発症予測に関して、単位時間当たりの発症患者の増加率が、いわば時間の逆数に関して二次式で表されるモデルを考え、それに基づいてモデルの未知母数を推定する。

次に、散在的集団発生に関して、発症日のパターンは逆ガウス型分布で記述可能であり、

同一の原因物質を摂取した集団の発症パターンにおける各未知母数は、必然的に同じ値を取ることを仮定している。そこで標本集団では構成が異なり（年齢分布）必然的に、発生分布の推定した未知パラメータが異なることは自然である。

以上に基ついて二つの2母数逆カウス型分布の母数に関する検定方式の構築を行った。

この検定方式の課題は、感染した時刻と発現した時刻とかわかっていなければならぬ点である。しかし通常は感染した時刻はわからないと考えるのが自然である。そこで適当な時刻の起点から何日目に症状が発現したとの情報しかないときに、その適当な時刻の起点から何日目か感染日であると推定することが必要である。

確率変数 X を感染してから発現するまでの時間とする。適当な時刻の起点からの時間を δ とすれば、適当な起点からの時刻 $\delta + X$ （定数とする）の実現値が観測値となる。無論 δ は未知である。従つて、 X が逆カウス型分布に従っていると仮定しても、適当な起点からの発症の時刻である $D = \delta + X$ は逆カウス型分布には従わない。ここで原点未知の3母数逆カウス型分布の利用が必要である。

構築した手法の問題点として以下があげられる。

- (1) 予測可能になるまでに最低4周期の時間が必要になる。
- (2) 曝露時の時点が未知であるケースには提供できない。
- (3) テータがばらついているものにも適用が難しい。

実用上においては、以下のような対策を施すことにより以上の課題が解決できると考える。

- (1) なるべく階級の幅を小さくして、正確に初期発症時の度数分布表を作成する。出来れば、6時間間隔か12時間間隔が望ましい。
- (2) 実用上は、付加情報から探索的に曝露時の予測値を決めて適用する。
- (3) 迅速な疫学調査を行い reporting delay や batch reporting によるデータのタイムラグを減らす。また、より重要なことは、この方法は、差分を利用しているの、度数分布の変動に関して非常に敏感に反応し推定値が適当でない値を求めるケースも生じる。そこで、統計的な性質は保証できないか、

度数分布を平滑化し、各時点の予測値を用いことを薦める。

D 結論

本研究では当面の統計的に解決すべき問題として、様々な問題を検討し、研究協力者との議論も踏まえ最終的には、以下の3つの問題を取り上げた。

- (1) 患者発生パターンの立ち上がり日の推定問題
- (2) 2母数逆カウス型分布に従う2つの母集団における母平均、母変動係数に関する検定
- (3) 微分方程式を利用した発生パターン曲線の算出

以上の問題を取り上げた理由は、それぞれが最短発症日の推定、散在的集団発生 (diffuse outbreak) への対処、集団食中毒の規模の早期予測を解く鍵となる統計解析手法であるからである。

0-157 を原因とする食中毒の発生パターンは、逆カウス型分布を用いて記述可能であると仮定して、統計理論の構築を進めた。この仮定は、単なる分布の取り扱い易さや過去において誰かこの分布を使っていたからといった理由ではなく、患者発生メカニズムと分布自身が持つメカニズムが対応できる点か様々な統計的問題を考える場合において非常に重要である。

E 研究発表

1 口頭発表

金藤浩司、岩瀬 晃盛

Estimation for a scale parameter with known coefficient of variation、第67回日本統計学会講演報告集、pp 75-76 1999 7 29

2 学会誌発表

R P Galapate, A U Baes, K Ito, K Iwase and M Okada (1999)、Trihalomethane formation potential prediction using some chemical functional groups and bulk parameters、Water Res Vol 33, No 11, pp 2555-2560

分担研究報告書

統計情報の有効利用に関する研究

分担研究者 馬場康維 統計数理研究所

研究要旨 地理情報を用いた既存の統計データの有効利用について研究した。

A 研究目的

コンピュータの高速化、大容量化、さらにネットワークの普及により、様々な複数のデータを結合利用することが簡単になっている。そこで、健康危機情報に関して、その他の情報の有効活用について研究する。

B 研究方法

地理情報システム (GIS) の利用は、現在非常に簡単になっている。そこで、GIS を一般的な視覚化の観点ではなく、データ結合の利用手段として用いることを検討した。データ結合の可能性の検討として、各省庁等が公開しているデータについての一部調査等を行った。

C 研究結果

現在各省庁は、様々な有効な統計情報をホームページ等で公開している。厚生省においても、0-157 の発生情報を緊急情報として公開している。例えば、月別都道府県別発生状況には、その発生事例ごとに届け出保健所まで公開されているので、GIS によってその他の地域特性とのマッチングが容易である。

D 結論

集団食中毒の発生は、日本各地で起こっている。その場合、各地の保健所と国立感染症研究所との連携のみならず、各地の保健所と地方の大学との連携が重要である。そこでの協力の内容としては、統計科学としての疫学調査への関与があげられる。その場合、首都圏の研究者だけでなく、発生地域に近い統計学者の協力が必要となる。そこで調査・研究等に協力可能な統計学者のデータやその内容に関するデータの把握、情報の結合、および人的ネットワークの整備が望まれる。

E 研究発表

1 口頭発表

馬場康維、地理情報を用いた統計データの利用、第 67 回日本統計学会講演報告集, pp 273-274, 1999 7 30

Y Baba, The present and future of statistical education, International Statistical Forum, Korea, 1999 9 17

中村好宏、馬場康維、大隅昇、損失関数最小化による射影と変数の分類、ISM Symposium, Recent Advances In Statistical Research and Data Analysis, Tokyo, 2000 3 22

參考資料

健康危機関連統計の高度処理に関する研究

(研究課題番号 H10-統計-005)

平成11年度厚生科学研究費補助金

(統計情報高度利用総合研究事業)

研究成果報告書

平成12年3月

研究代表者 金 藤 浩 司

(統計数理研究所・統計科学情報センター)

まえがき

従来の統計情報の公開方法のみならず、緊急時において国民・地方自治体等から、迅速でかつ多面的な統計情報提供の要望がある。本研究では、腸管出血性大腸菌O-157（以下ではO-157と省略する。）を原因とする食中毒の発生状況に関して厚生省が保有する統計情報を処理し、国民・地方自治体等に質の高い統計情報を伝達する基礎となる統計解析システム構築に向けた基礎的研究を目的とする。

厚生省は、平成8年より健康危機関連情報の一つとしてO-157に関する発生状況をホームページ上で公開している。これは、平成8年に大阪府で発生した学校給食に関わる集団食中毒が国民に不安を与え、厚生省が国民からの情報提供の要望に答えたものである。他機関の健康危機関連情報の公開に関しては、国立感染症研究所、国立医薬品食品研究所、首相官邸、農林水産省においても関連情報を国民に伝達するためにホームページが開設され今日に至っている。

現時点で厚生省のホームページで得られる食中毒に関する情報としては、O-157のみならずその他の病因物質別の発生状況が公開され、O-157に関する特別の集計ではなくなっている。このことは、一時期のような集団発生はおさまったように行政等が感じている表れかもしれない。しかし、大阪府で発生したものと同程度の学校給食が原因となる集団食中毒の可能性は否定できない。例えば、日本体育・学校健康センターの資料（平成10年5月1日現在）によると、共同調理場数あたり5000人以上の児童生徒の給食を作っている調理場は全国に237カ所ある（共同調理場数はここ10年あまりは約2700カ所で推移している）。

本研究では健康危機関連統計を処理する統計手法の構築とともに、健康危機関連情報の時間・空間的解析等を行うことにより、疾患の経年変化や地域特性を解明する基となる可視化システムの検討を行った。また、これにより予測を含め新たな事象が発生した場合への迅速な対応が可能となり、質の高い統計情報が提供可能となる。

健康危機に関する状況は、O-157のみならずますます重要な政策上の問題となっ
てきている。そこで、本研究が厚生行政の政策立案や厚生研究の一助になれば幸いである。

平成12年3月

研究代表者 金藤 浩司

研究組織

- 研究代表者 金藤浩司（統計数理研究所統計科学情報センター助教授）
- 研究分担者 越智義道（大分大学工学部助教授）
- 岩瀬晃盛（広島大学工学部教授）平成11年4月1日より
- 馬場康維（統計数理研究所統計科学情報センター教授）平成11年4月1日より
- 研究協力者 渡辺治雄（国立感染症研究所細菌部長）
- 谷口清州（国立感染症研究所感染症情報センター室長）
- 瀬上清貴（厚生省大臣官房統計情報部保健統計室長）
- 南 弘征（北海道大学情報メディア教育研究総合センター助教授）
- 清水信夫（北海道大学大学院工学研究科大学院生）
- 高橋志乃（大分大学工学部教務員）
- 井上顕隆（大分大学大学院工学研究科大学院生）
- 吉田 烈（大分大学大学院工学研究科大学院生）

研究経費

平成10年度	1,700千円
平成11年度	5,000千円
合計	6,700千円

目次

	(執筆分担者)	ページ
第1章 総論	金藤浩司	1
第2章 健康危機関連統計に関する統計理論の構築	岩瀬晃盛 金藤浩司	5
第3章 統計解析システムの実装に関する検討	金藤浩司	21
第4章 データの視覚化に関する検討	越智義道	35
第5章 地理情報を用いた統計データの利用	馬場康維	103

第1章 総論

1. はじめに

平成8年に大阪府で発生した学校給食を原因とする集団食中毒において、病因物質が病原性大腸菌O-157であったことは記憶に新しい。O-157を病因物質とする食中毒は国立感染症研究所O-157関係ページによると、1982年アメリカのオレゴン州とミシガン州で発生したハンパーカーを原因とする食中毒で、O157 H7型菌が原因菌である事が初めて報告され、日本では1990年10月に埼玉県浦和市で発生した事件において報告され現在に至っている。

これまでの健康危機管理に対する両国の対処に関する経緯をながめると、アメリカのCDCを中心とした迅速な対応に比べ、現時点の日本では相当の遅れがある。その理由としては、これまでの日本における国民および行政の健康危機管理に関する重要性の認識の薄さがあげられる。しかし、大阪府での集団発生により国民に生まれた健康危機管理意識の高まりとともに、行政への情報提供の気運が高まり、政府特に、厚生省、文部省、農林水産省等の関係省庁やそれに付随する研究所等からの情報発信が行われている。

さらに厚生行政に限ると、近年、厚生省、国立感染症研究所（感染症情報センター）、地方の保健所等からの食中毒に関する発生情報のデータ送信や、発生に対処する情報の伝達に関するネットワークが計画(Food Net)されていることは、日本における食中毒関連の健康危機管理に関して意義が大きい。そこでは、原因菌の解析、汚染ルートの解明、汚染物質の回収等の行政措置による感染の拡大防止、データベースの構築が計画されている。このようなネットワークの構築は、健康危機に関して迅速に対処するための基礎となる。しかし、システムが有効に効率よく動作するためには継続的で積極的な疫学調査が必要不可欠であることは言うまでもない。また、収集したデータをいかにして適切に処理し、より有効により効果的に利用するかといった問題が残されている。

以上のような状況を踏まえ、いかにして食中毒に関する健康危機関連情報の客観的な収集計画を策定し、実際に収集し、収集したデータを統計解析し、公開するかを検討していかなければならない。

そこで本研究では2つの基礎的研究目標をかかげた。

- (1) 健康危機統計に対する統計解析手法の構築に関する基礎的研究
- (2) O-157の発生状況に関わる可視化に関する検討

2. テータ

初めに、現状で利用可能なテータの検討を行った。基礎テータとしては、厚生省のホームページから入手可能な(であった)O-157情報を基に、平成9年、平成10年、平成11年の基礎テータの整理を行った。このテータは厚生省生活衛生局食品保健課が管理し速報としてホームページに掲載されている。速報性を優先するためこのデータをそのまま統計解析に用いる場合には問題がある。その詳細は第4章で述べるが、一例をあけると、数人のテータがまとめて掲載されていたり、発症日の情報が欠損していたりする。しかし、当面、誰でも利用できる点においてこのデータは非常に意義がある。

また、食中毒に関するその他の統計として伝染病統計がある。これは総務庁所管の統計法に基づく届出統計であるため個票の使用に関しては目的外申請が必要となり、研究の基礎的資料としては重要であるが簡単に誰でもが利用できる状況ではない。

そこで、本研究で検討・構築する統計手法と可視化に関しては現状で得られるデータに限定して考える部分となんらかの方法で将来得ることが可能であるデータに関する部分を明確に記述する必要がある。

3 統計解析手法の基礎的研究

本研究で当面の統計的に解決すべき問題として様々な問題を検討し、研究協力者との議論も踏まえ最終的には以下の3つの問題を取り上げた。

- (1) 患者発生パターンの立ち上がり日の推定問題
- (2) 2母数逆カウス型分布に従う2つの母集団における母平均、母変動係数に関する検定
- (3) 微分方程式を利用した発生パターン曲線の算出

以上の問題を取り上げた理由は、それぞれ最短発症日の推定、散在的集団発生(diffuse outbreak)への対処、集団食中毒の規模の早期予測を解く鍵となる統計解析手法であるからである。

また、O-157を原因とする食中毒の発生パターンは逆カウス型分布を用い

て記述可能であると仮定して統計理論の構築を進めた。この仮定は単なる分布の取り扱い易さや過去において誰かがこの分布を使っていたからといった理由ではなく、患者発生のメカニズムと分布自身を持つメカニズムが対応できる点が様々な統計的問題を考える場合において非常に重要である。このことについては詳細を第2章で述べる。

また、対数正規分布を用いた研究やその他の寿命分布を用いた研究もあるが、本研究では先に述べた理由により患者発症のパターンと解析に用いる分布の形が似ていればよいという立場は取らない。

同時に、第2章で考えている問題の範囲は、地域を限定し同一の原因物資に曝露された集団の時間に関する情報である。そのため、日本全体での患者発生パターンや傾向を議論するものではない。また、後者の議論の道具は第4章で検討している。

4. 視覚化に関する基礎的研究

0-157のデータは、時間情報のみならず地理的発生情報を含んでいる。そこで、本研究では地理的情報を含んだデータをグラフィカル表示することにより、経時変化・地域性などの特性を考慮に入れて疾患の発生状況を分析し、情報を提供することを目的とするシステムの構想について議論する。

第2章ではデータの視覚化について以下の2つの観点から検討を行った。

- (1) Unix 環境下での動的なグラフィカル・データ解析システムについて0-157関連情報を表示するための位置情報と時間情報の表示に関する検討
- (2) GISシステムを採用して、0-157関連情報の提示可能性についての検討

以上の観点を対象とした理由を要約すると、健康危機関連情報では従来型の単なる要約統計の公開ばかりではなく、即時的で多角的な情報の提供が求められる。そこで、多次元データの視覚化表現により、その位置や時間における集積性を同定することの可能性の検討が必要である。

また、探索的な分析ではデータを視覚化表現し、解析者の視覚的な情報処理能力を利用することが有効であることが知られている。また本研究の対象となるような地理的な情報を含むデータの分析では地理的特性やその地理的関連性を把握する上でその情報を視覚的に表現することが不可欠な要素となる。

5. 統計情報の利用

集団食中毒の発生は日本各地で起こっている。その場合、各地の保健所と国立感染症研究所との連携のみならず統計科学としての調査が必要である。またその場合、首都圏の研究者だけでなく、発生地域に近い統計学者の協力が必要となる。そこで研究に協力可能な統計学者のデータやその内容に関するデータの把握が必要である。

6 まとめ

これまで述べたことを中心に、本研究では様々な問題について研究協力者を含めた議論を行ってきた。その結果は以下のようにまとめられる。

統計解析手法に望まれる点として、O-157は原因物質を摂取してから発症までの期間が他の食中毒に比べ長い場合何らかの統計的な処理手法（推定、予測）が必要になる。そこで原因物質の候補となる物質を探すために原因物質を摂取した日の候補をまず探さなければならない。また、個体差によって発症日が異なるため、早期の対策を施すために得られたデータで集団食中毒の規模の推定が必要になる。また、原因食材は全国に流通している可能性があるため、散在的集団発生を引き起こす可能性が高い、そこで正確な判断は菌のDNA鑑定等が必要であるが、事前に統計的な手法としての母集団の比較を行うことによってより効率的な散在的集団発生への判断が下される。

ただし、現状のO-157データには、

- (1) 発症日が同定できない個体が多く存在する。
- (2) 集団としてのデータが確定されるまでに時間がかかる。

等の解決すべき点がある。そのためにもさらなる統計手法の構築が必要である。また、基礎情報としてより詳細なデータが個人のプライバシーを守られた形で迅速に利用できる環境が必要である。

また、以上のような統計解析によって得られた結果を分かり易く表現するためには、多面的な表示に基づく情報の提供が必要である。そのためには、データの視覚化が必要であり、これは最終的に得られた結果の公開方法において効果的であるのみならず解析途中における探索的データ解析のためにも重要である。

また、本研究をさらに進めることによって、厚生省が構築しているFoodNetとの相互補完的な関係が可能である。

第2章 健康危機関連統計に関する統計理論の構築

1 はじめに

本研究では、厚生省のホームページから得られたデータや学会誌等で公開されているデータのみならず、集団食中毒の発生時に国立感染症研究所、各保健所にデータが迅速に集められた場合に、どのような統計解析が可能であるか検討している。

O-157に関して得られるデータには、各個体でのO-157の発症日、保健所探知日、患者発生日等の時間に関する情報が含まれている。しかし、必ずしもすべての発症者に対して正確なそれらの情報が得られないことは、解析の妨げとなっている。厳密には、データの精度に関係した統計的手法の研究を進める必要があるが、これらの研究についてはこれからの課題である。ここでは、これらの時間に関する統計情報についての統計解析手法に関する基礎的研究を行った。

一般に、O-157を原因菌とする集団食中毒の発生は、ほぼ同時期に原因食材を摂取した患者が、平均して数日後に食中毒の症状を訴えることによって確認される。ここで公衆衛生上問題となる点は、O-157を原因とする食中毒の確認までの期間に個体差があり、他の原因菌を原因とする食中毒に比べ食中毒の発生の確定までの時間が長い点である。このことが原因食材の推定や確定を難しくしている。これは、相当数の児童生徒が同じ食材を摂取する学校給食を原因とした場合においても、最終的な原因食材の確定を困難にしている要因である。

以上のような状況において、最終的な医学・細菌学・疫学的検知からの原因食材等の確定は必要であるが、人的・物的資源の問題により、各地で発生している集団食中毒への迅速な対応には至っていないと思われる。そこで現状の行政システムで得られるデータを効果的に利用するため、探索的・仮説検証的な手順として、統計解析手法が有効であり、同時に統計情報資源の有効活用の意味からも重要である。

本研究で焦点を絞った統計解析上の課題は以下の3点である。

- (1) 患者発生パターンの立ち上がり日の推定問題
- (2) 2母数逆ガウス型分布に従う2つの母集団における母平均、母変動係数に関する検定
- (3) 微分方程式を利用した発生パターン曲線の算出

これらの項目は、あくまでも探索的・仮説検証的な手順であり、これ以降のより詳しい疫学調査や細菌学におけるDNA検査を効率的に行うための有効な道具となるもの

てであると考えている。また、これらの統計解析を基に、より絞り込んだ形でその他の確定作業が行われることが望ましい。同時に、迅速で集中的な疫学調査に基づくデータ収集が行われることにより統計解析の精度が上がることを期待できる。

2 逆ガウス型分布

本章の研究では、逆ガウス型分布をすべての統計手法の前提としている。初めに、患者の発症日のパターンと逆ガウス型分布の関係を述べる。

初めに解決しなければならない課題として、原因物質が体内に入った時点の推定がある。なぜなら本章で考察する逆ガウス型分布等の寿命分布を用いる統計解析手法の構築においては、原因物質が体内に入り発症するまでの時間のデータが必要である。しかし、0-157の時間に関する情報を統計処理する場合、原因食材が確定しにくい（原因食材を摂取した時点の確定が難しい）。そのため、発症までの正確な時間情報が得られず、得られるのは発症日の情報のみの場合が多い。そこで、本研究の枠組みで必要である原因食材を摂取した時点の推定に関する研究の必要性が生まれる。また、この課題は原因食材の推定においても重要である。そこで、これに関する手法について考察する。

また、解析するデータの発症パターンが逆ガウス型分布を前提にそれ以降の解析を進めてよいかどうかといった問題がある。そのために、逆ガウス型分布確率紙による逆ガウス型分布の有効性に関する検討を行った。

2 1 患者の発症パターンと逆ガウス型分布

逆ガウス型分布は、正のドリフトを持つウィーナー確率過程のある値への初期到達時間分布として導出される。このような分布自身が有するメカニズムが、本研究では重要な意味を持っている。

初めに、0-157データにおける患者の発症パターンのメカニズムを考える。そこでは、独立な個体が数多くいるところに共通の要因で病原菌が摂取され感染したとしたときに、個体の代謝能力やその他生活環境等の相違により発症までに要する時間に個体差が生じる。ここでは、感染して直後に発症する人をも確率が極めて低いにせよ想定している。

もしも決定論的に考えて感染してからの時間に比例して病原菌が増殖して、ある閾値に達したら発症すると考え、完全に同じ（いわばクローン人間）集団であると考え全く同じ生活環境であるとすれば、感染者は全く同じ時刻に発症する。この時刻が2母数逆ガウス型分布で言えば平均値に相当する。

実際には、個体は同じでなく生活環境も異なるため発症までの時間に変動が生じ、発症時間分布が逆ガウス型分布に従うと考えるのが妥当である。次に、2母数の逆ガウス型分布の定義とその性質について述べる。

2 2 2母数逆ガウス型分布

形式的には、未知母数の入れ方によって異なる2母数逆ガウス型分布の表現形式が見られる。ここでは、以下のような確率素分 $f(x)dx$ に従う確率変数 X の従う分布で2母数逆ガウス型分布 $IG(m, c^2)$ を定義する。

$$f(x)dx = \frac{1}{\sqrt{2\pi}c} \left(\frac{x}{m}\right)^{\frac{3}{2}} \exp\left[-\frac{1}{2c^2} \left(\sqrt{\frac{x}{m}} - \sqrt{\frac{m}{x}}\right)^2\right] \frac{dx}{m}$$

ここで、 $0 < x < \infty, 0 < m < \infty, 0 < c < \infty$ である。

以上のように定義された2母数ガウス型分布にデータが適用できるかどうかについては、適合度検定等の利用が可能である。しかし、ここでは視覚的な観点から確率紙による検討を行う。一般に、確率紙としては、正規確率紙、ワイブル確率紙が有名である。一方、2母数逆ガウス型分布に関しては、岩瀬他(1991)があり、次節では、その作成手順のアルゴリズムについての概略を示す。

2 3 2母数逆ガウス型分布確率紙の構成

岩瀬他(1991)は、逆ガウス型確率紙の作成を試み、その検討を行っている。理論的な構成法については、本節では省略し、実際に作成するアルゴリズムの概略についてのみ要約して述べる。初めに、以下のことを仮定する。

仮定

同一の母集団 $IG(\mu, c^2)$ から無作為に大きさ n の標本 (x_1, \dots, x_n) が抽出される。

2 3 1 確率紙作成のアルゴリズム

step 0 大きさ n の標本から、標本算術平均 $\bar{x}_A = \frac{1}{n} \sum_{i=1}^n x_i$ を計算する。

step 1 各 x_i について $x_i < \bar{x}_A$ の場合、 $k_i = \frac{\bar{x}_A}{x_i}$ を計算し、 $x_i \geq \bar{x}_A$ の場合、 $k_i = \frac{x_i}{\bar{x}_A}$ を計算する。