

はじめに

1.1 研究の背景

1995 年頃より爆発的にブームを引き起こしたインターネットに代表されるように、ここ数年の情報通信分野の発展には目覚ましいものがある。このインターネットとともに普及したのが Integrated Services Digital Network (ISDN) である。そして ISDN よりも高速大容量の情報通信サービスを提供するものとして、Asynchronous Transfer Mode (ATM: 非同期転送モード) 網や Asymmetric Digital Subscriber Line (ADSL: 非対称デジタル加入者回線) 網に代表されるような x Digital Subscriber Line (xDSL) [94]が考えられている。ATM 網は主に大都市間の基幹網に用いられている。また、ADSL 網を用いたインターネットサービスが 1999 年 9 月より川中島町 (長野県) で開始され、都区内でも年度内にサービスが始まる予定である。このような網を実装することにより、インターネットのサービスはもちろんのこと、将来的には家にいながら好きなときに好きな番組を見ることができる動画配信サービス Video On Demand (VOD) 等のサービスを受けることができるようになることが予想される。

このような高速情報通信網を流れるデータ、音声、圧縮動画等のトラフィックは、従来の電話やパケット交換機のそれとは異なり、自己相似性と呼ばれる性質をもつといわれている。自己相似 (フラクタル) 性は、幾何学において有名である。幾何学的な解釈は、図形の任意の一部分をを拡大したときもとの図形と同一のものになっているという性質である。この代表的な例として、Koch 曲線がある (図 1.1)。また、幾何学的なものだけではなく、自然界にも自己相似性をもつといわれるものが数多く存在する。例えば、海岸線の形や木の構造、雲の形等は自己相似性をもつといわれている [87]。通信網上の時系列データにおける自己相似性は長記憶性と呼ばれる性質をもつ。長記憶性とは、通信トラフィック量を時系列データ $\{X_n; n = 1, 2, \dots\}$ (単位時間あたりのバイト数等) として見たとき、その自己相関係数 $\rho(k) = \text{cov}(X_n, X_{n+k}) / \text{var}[X_n]$ が、 k が増えるとき $k^{-\beta}$ ($0 < \beta < 1$) に従って減少するとい

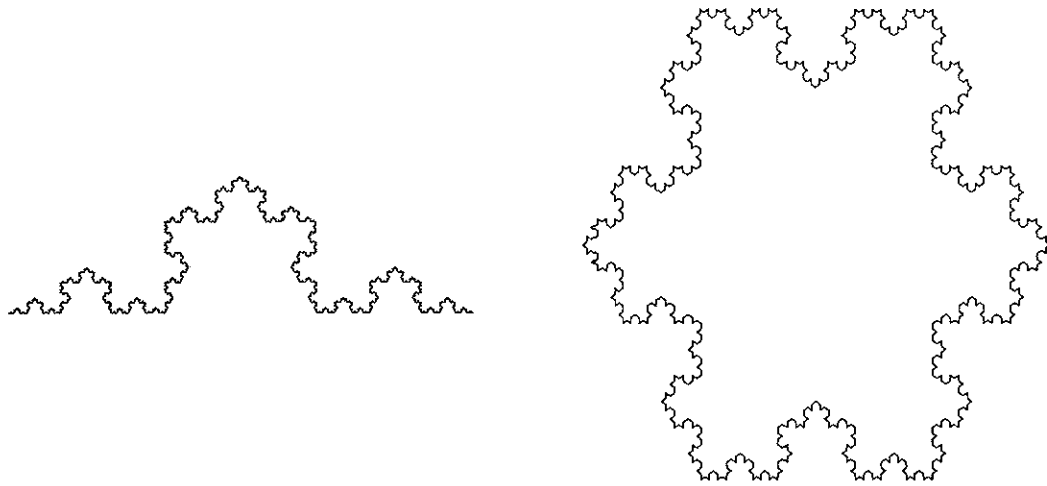


図 1.1: (左): Koch 曲線, (右): Koch の雪片.

う性質である. 通信網上の時系列データに関する自己相似性の例は, イーサネット Local Area Network (LAN: 域内通信網) [55, 56, 98, 100]をはじめ, Wide Area Network (WAN: 広域網) [67, 68], ATM 網 [34, 46, 93] で観測されている. また, 本論文で扱っている Variable Bit Rate (VBR: 可変ビットレート) で符号化されたデータにおいても自己相似性が観測されている[3, 15, 31, 35, 36, 39, 44, 48, 49, 50, 53, 54, 59].

それ以外の例に関しては, [99]に示されている. ここにあげた論文のほとんどで, Poisson 型のモデルがもつ無記憶性とは全く違う性質が現れていると報告されている. 従って, 従来の交換機等のシステム設計に用いられていた Poisson 過程に基づく理論をそのまま用いることができないことが予想される. そこで, Poisson 型のモデルに代わるモデルを作る研究も多数報告されている. 例えば, LAN のトラヒックに対する [62, 70], 無線網 に対して Fractional Brownian Motion (FBM) を用いた[63], WAN に対する [67] 等等である. 特に, 動画像に対するモデル化が盛んに行われている[15, 31, 39, 34, 46, 47, 59]. しかしながら, Poisson 型のように解析的に扱いやすく, かつ, 汎用性に優れているモデルは研究段階である. 従って, このような高速網を流れるトラヒックの特性を研究し, そのトラヒックをモデル化することは重要である.

このように, 最近の情報通信網のトラヒックは自己相似性をもつといわれているが, 従来はこのようなトラヒックをモデル化する際に, 待ち行列モデルを用いていた. その最も基本になる待ち行列モデルが $M/G/1$ 待ち行列である. 通常, 待ち行列システムにおいて最も興味があるのは, 客がサービスを受けるまでどの程度待つかという待ち時間である. $M/G/1$ 待ち行列において, 平均待ち時間はサービスの順番によらず同じである. しかしながら, 2 次以上の高次モーメントはサービスの順番に依存する. 基本的なサービス規律として, First-Come First-Served (FCFS: 先着順サービス) 規律, Last-Come First-Served (LCFS: 後着順サービス) 規律, Random Order of Service (ROS: ランダム順サービス) 規律の 3 つが考えられる.

先着順サービスは、日常生活において、例えば、銀行の窓口で我々が受けるような到着順サービスである。一方、後着順サービスは、コンピュータのスタック等で用いられている。ランダム順サービス規律とは、これら 2 つのサービス規律とは異なり、サービスを待っている人の中から無作為に等確率で次にサービスされる人を選ぶサービス規律である。

更に、サーバの数を 1 つのままにして、待ち行列の数を増やしたシステムがある。その 1 つがポーリングシステムである。ポーリングシステムは、1 つのサーバが N 個の待ち行列を順番に回ってサービスをするシステムである。1 つの資源を順番に割り当てることによって、多くのユーザが資源を共有できるようになる。ポーリングシステムは、通信システムや生産システムに応用される。通信システムにおいては、高速 LAN のためのプロトコルである FDDI や、トークンリング LAN 等の性能評価を行うに際し、よく用いられている。

このように、既実装されているシステムにおいて、未だ明示的な結果が与えられていない問題を解くことも重要であると思われる。

1.2 情報通信網のトラヒックの自己相似性に関する研究

情報通信網のトラヒックは、以前より Markov 的ではないといわれていたが、実際、明示的に解釈できなかった。その問題に 1 つの解釈を与えたのが Leland ら [55, 56] である。それが自己相似性と呼ばれる性質である。自己相似性と呼ばれる性質は、単一ソースの場合にバースト性を示すだけでなく、いくつかのソースを重ね合わせた場合にもバースト性を示す。これは Poisson 過程とは全く異なる性質である。このような性質は、インターネットのおける、Teletype Network (TELNET) や File Transfer Protocol (FTP) 等で TCP セッションの確立後に通信網状を流れるデータ等にも見られる [68]。また、将来的に情報通信網を流れることが予想される画像データでも自己相似性が見られる (第 1.1 節)。このようなデータが流れる情報通信網上の交換機やバッファ、回線容量等を設計するときに、自己相似性をもつトラヒックを無視することはできないと思われる。

本節では、情報通信網のトラヒックの自己相似性に関する研究の概要を説明する。第 1.2.1 節は、情報通信網のトラヒックの自己相似性に関する先行研究を示す。第 1.2.2 節では、自己相似性についての定義と自己相似性を検出する方法を述べる。第 1.2.3 節は、自己相似性をもつ時系列データを生成する方法を説明する。

1.2.1 過去の研究

本節では、高速情報通信網のトラヒックが自己相似性をもつといわれるようになった文献をいくつか紹介する。以前より、データ通信トラヒックは Poisson 過程とは異なる性質をもつといわれていたが、その違いは明示的に示されてはいなかった。それが自己相似性をもつと初めて報告したのが、イーサネットの実測トラヒックデータからそのことを示した Leland [55, 56] や Willinger [98] である。

Leland らは、Bellcore Morris Research and Engineering Center (MRE) において、イーサネット LAN のパケット数を、連続した 20 時間から 40 時間に渡って測定した [55, 56]. これらのデータは、異なる社内 LAN 上で、異なる期間 (1989 年 8 月, 1989 年 10 月, 1990 年 1 月, 1992 年 2 月) に渡って測定されている。これらのデータのうち、1989 年の 8 月と 10 月に取られたものは、社内 LAN 内間のトラフィックである、つまり同じ職場内のホストどうしのトラフィックを測定したものである。これらは主にリモートログイン、電子メール、サーバから端末へのファイルサービスのための Network File System (NFS) プロトコルのような Internet Protocol (IP) である。1990 年のデータは 2 つの研究所間に監視システムを置き、その間を流れるトラフィックを取ったものである。そしてこの 2 つの研究所を結ぶ回線は外部のインターネットへも繋がれている。このトラフィックデータはホスト間とルータ間のものが混在している。そのため、このトラフィックにはいくつかのトラフィックが混在している。1992 年のデータは MRE のイーサネットの基幹回線を流れるトラフィックを取った。これは、MRE 内の研究所間を行き交うすべてのトラフィック、他の Bellcore の建物から MRE へ向けられたトラフィック、Bellcore の外へ向けられたすべてのトラフィックである。これは主にルータ間のトラフィックを測定したものである。

1989 年の 8 月のデータでは、観測時間を過渡期、通常期、閑散期の 3 つに分け、それぞれ 360,000 個の測定値をもつようにしている。通常期のデータを用いて R/S 解析、分散解析、及びピリオドグラムによる解析を行ったところ、自己相似パラメタ H の推定値は 0.8 くらいであることが分かった。更に、集積過程の集積レベルが $m = 1, 5, 10, 50, 100, 500, 1000$ の場合の各解析の推定値をグラフに示したところ、集積過程の分散に対して最小二乗法によって得られた傾きが 0.8 ではほぼ直線になっている。このことから 1989 年 8 月の通常期のデータは $m \leq 1000$ の範囲では厳密な 2 次の自己相似性になっていることが分かるといっている。過渡期、閑散期におけるデータも同様に自己相似性を示していた。

次に、Leland らは 4 年間に渡るイーサネット LAN のトラフィックについて自己相似パラメタ H を推定した [55, 56]. これは各データの通常期のデータに関して最尤推定を行ったものである。最尤推定の結果、集積レベル $m = 100, 200, \dots, 2000$ における自己相似パラメタの推定値とその 95 % 信頼区間を示し、その信頼区間の間に、 R/S 解析と分散解析によって推定された自己相似パラメタ H が入っていることを示した。更に、4 年間に渡って測定したすべてのデータについて、過渡期、通常期、閑散期に分け、そのそれぞれについて、パケット数とバイト数ごとに R/S 解析、分散解析、最尤推定とその 95 % 信頼区間を示した。パケット数に関する結果から、過渡期のデータの方が通常期に比べ、自己相似パラメタ H の値が大きいことが分かった。また、従来の Poisson のようなデータでは、ソース (イーサネットに接続してデータをやりとりするユーザー) の数が増えるとトラフィック全体がなめらかになるのだが、イーサネットのトラフィックでは逆に、なめらかになるどころか、よりバースト的になった。また、バイト数に関する結果からは、ホストの計算能力の向上によってパケットが自己相似性の形に近づくようなビット率で発生してしまい、その結果、自己相似パラメタ H の値を大きくさせてしまったことが読みとれた。閑散期における信頼区間が他の 2 つの期間に比べ

広がっている傾向が見られたが、これは閑散期のトラヒックは、厳密な自己相似性というよりは、むしろ漸近的な自己相似性をもつからであると説明している。

今まで、Bellcore 中のトラヒックつまり内部網のトラヒックについて調べてきた。しかし、Bellcore の中から Bellcore 以外のところへ向かうトラヒックや、逆に、Bellcore の外から Bellcore の中へ入ってくるトラヒックもある。文献 [55, 56] ではこのようなトラヒックについても考察している。このようなトラヒックは、1990 年と 1992 年に測定したデータである。今までと同様に、測定データを内部網で考えた過渡期、通常期、閑散期に分けそれぞれの期間でのパケット数とバイト数を測定した。その結果、内部網のトラヒックに対して、外部網のトラヒックは、閑散期でおよそ 1-10 %、過渡期で 2-25 %、通常期では 56 %、までを占めていた。内部網と外部網のトラヒックの違いは、サーバとワークステーション間の NFS のトラヒックの有無である。自己相似性に関する外部網のトラヒックの過渡期と通常期の解析の結果は、内部網の結果と違いはなかった。例として、1990 年の過渡期のデータでは、外部網のトラヒックは全体のトラヒックの 2 % に過ぎないけれども、最初の 30 分間のデータでは $H = 0.82$ 、次の 30 分間では $H = 0.94$ と推定された。対応する内部網のトラヒックの解析結果では $H = 0.98$ であった。しかし、閑散期のデータでは、顕著な相違が見られた。内部網のトラヒックデータでは $H \approx 0.7$ であったのに対し、外部網のトラヒックデータでは $H \approx 0.55$ であった。

文献 [55, 56] と同じデータを用い、主に自己相似性の解析方法やモデル化に力点を置いたのが Willinger らの [98] である。この論文では、Beran ら [6] によって示された大きなデータに対して素早く自己相似パラメタ H を推定する方法と、文献 [55] に述べられていた、1 次の autoregressive 過程 (AR(1)) を多重化させたり renewal reward 過程によって測定データに似た形を表すデータ列を作り出す方法を、紹介している。

Beran ら [6] による方法は、次のようなものである。 n 個の要素からなるもとの時系列データを、 i 個ずつの部分列に分解し、その部分列ごとに最尤推定法のような既存の方法で自己相似パラメタ H を推定する。更にその推定値の平均を取る。このようにして求められた自己相似パラメタ H の推定値は $i/n \rightarrow \gamma$ になるように $n \rightarrow \infty$ とすると、 $n^{1/2}(\hat{H} - H)$ はもとの時系列データに対する Whittle の推定値の漸近分布と等しくなる。この方法は、非常に長い時系列データに対して、適切な環境が与えられれば、実時間レベルでのパラメタ推定を可能にする。この方法は、非常に大きなデータに対して、コンピュータでパラメタを推定できるということ以外に、時系列データ全体に渡って自己相似パラメタ H が一定であるかどうかを調べる方法を与えている。これはまず、上述のように分けた $k = [n/i]$ 個の部分列の推定値 $\hat{H}_1, \dots, \hat{H}_k$ に対して $T_{1,2,\dots,k} = \sum_{j=1}^{k+1} (\hat{H}_j - \hat{H})^2 [\hat{\sigma}_j^2 i]^{-1}$ という統計量を定義する。ここで、 $\hat{\sigma}_j^2 i$ は Whittle の推定量に対する中心極限定理の一般化によって与えられる。帰無仮説を “もとの時系列データに対して $H = H_j (j = 1, \dots, k)$ が成り立つ” とし、対立仮説を “ $H_j \neq H_i$ となるような $j \neq i$ が少なくとも 1 つ存在する” とし、仮説検定を行う。このとき、帰無仮説のもとで、 $T_{1,2,\dots,k}$ が漸近的に自由度 $k-1$ の χ^2 分布に従っているということを示した。よって、 $\chi_{k-1;\alpha}^2$ を自由度 $k-1$ の χ^2 分布の $(1-\alpha)$ 分位点として、もし

$T_{1,2,\dots,k} > \chi_{k-1;\alpha}^2$ が成り立つならば、有意水準 α で帰無仮説が棄却される。この方法をイーサネットのトラフィックデータに適用させたところ、部分列の数が2つの場合以外では、帰無仮説は棄却されなかった。これは、もとの時系列を3つ以上に分割したとき、各部分列の推定値をもとの時系列の部分列の推定値としてよいということの意味している。この方法により、計算量をかなり減らすことができる。

イーサネット同様、VBR 符号化された圧縮動画像等でも自己相似性が観測されている [5, 15, 31, 35, 36, 39, 44, 48, 49, 50, 53, 54, 59]。その中から、文献 [5] について簡単にその内容を記す。

Beran [5] はテレビ会議、テレビ電話、ニュース、スポーツ番組、教育番組、ドキュメンタリー、人気のあるテレビドラマ、映画等の動画像 (約 15 秒 - 2 時間) をデータとして用いている。符号化方式は H.261 を VBR 符号化させた符号化方式等を用いているが、ほとんどが公開していない符号化方式を用いている。文献 [5] では、これらのデータに対して、分散解析、 R/S 解析、及びピリオドグラムに基づく Whittle の最尤推定法により、長記憶性の有無を調べている。これらの方法により、VBR 符号化されたデータが、符号化方式や画像ソースによって異なる自己相似パラメタ H の値を取るかどうかを調べた。符号化方式の異なる、2つのテレビ会議の画像を分散解析した結果、自己相似パラメタ H の推定値はほぼ同じであった。また、このビデオ会議の画像の一方によく合うように Markov 連鎖を用いたシミュレーションによって作られたデータに対しては短記憶性の特性が現れた。つまりこのモデル化では、長記憶性を表すことができないことを分かった。分散解析では、集積レベル m の小さい範囲は過渡的なため無視する。無視する m の値が 200-500 くらいまではかなり変動しているが、 m が 500 を越えると安定し始めることが分かった。 R/S 解析によれば、2つのテレビ電話の画像については Hurst パラメタの値はほぼ同じであったが、ニュース番組を VBR 符号化したものと比較するとニュース番組の方が Hurst パラメタの値が大きかった。この違いは、シーンによる違いと Beran らは考えた。そして、Markov 連鎖を用いたシミュレーションによって得られたデータは R/S 解析でも短記憶性の性質を示した。

Beran らは次に、もとの時系列データからそれぞれ 1 (もとの時系列データそのもの)、10, 50, 100 個の部分列を作り、それらの平均を取ることににより新しい時系列データを作った。新しく作られた時系列データに対して、 R/S 解析を行った。例えば、もとの時系列データが $\{X_n; n = 1, 2, \dots\}$ でこのデータから 10 個の部分列を作ると、 $X_i^{(10)} = \{X_k; k = ((i-1)n)/10 + 1, ((i-1)n)/10 + 2, \dots, (in)/10\}$ となる。この平均を取った $X^{(10)} = (1/10) \sum_i X_i^{(10)}$ に対して、 R/S 解析を行った。その結果、ビデオ会議の VBR 画像では、この平滑化の影響がなく、もとの時系列データとほとんど同じ結果が得られた。しかし、Markov 連鎖を用いたシミュレーションによって得られたデータは、 $H = 0.5$ の漸近線の周りに R/S 統計量の対数の値がかたまってきた。このようにして短記憶性と長記憶性を区別することが可能である。

最後に、ピリオドグラムによる方法を用いて、データ解析を行っている。ピリオドグラムを用いる方法は、Willinger ら [98] が述べているように、Gauss 性による変動による荒さの

問題が生じてしまう。この問題を克服するために、Beran ら [5] は対数変換したデータに対してピリオドグラムを計算した。更に、もとの時系列データの他に、前に述べた平滑化したデータを対数変換したもののピリオドグラムも計算した。それらを比較すると、平滑化しないデータによる Hurst パラメタの推定値は、明らかに大きい値に偏ってしまったが、平滑化レベル (分割数) を増加させると推定値の偏りは減少し、分割数が約 200 ではほとんどなくなっている。各 VBR 画像データを対数変換したものに対して、ピリオドグラムと平滑化を組み合わせた方法で Hurst パラメタを推定した。その結果、Hurst パラメタの推定値 \hat{H} がほとんどが 95 % 信頼区間に 0.5 を含まないことから、これらの VBR 画像データは長記憶性をもつことが確認され、長記憶性の強さはシーンの動きに依存しているという結論に達している。

1.2.2 自己相似性とその解析方法

本論文の第 2 章では、高速情報通信網のトラヒックがもつといわれている自己相似過程と長記憶性について、既存の解析方法を説明する。

時系列データに関する自己相似過程は、ある離散時間確率過程 $\{Y_n; n = 1, 2, \dots\}$ に対して、任意の正の整数 c と実数 $H \in (0, 1)$ があり、 $c^{-H}Y_{cn}$ の分布ともとの Y_n の分布が等しくなる過程であると定義される。このとき、 H を自己相似パラメタと呼び、自己相似性の尺度を表す。自己相似パラメタ H をもつ自己相似過程 $\{Y_n; n = 1, 2, \dots\}$ の増分過程 $X_n = Y_n - Y_{n-1}$ の自己相関係数 $\rho(k) = \text{cov}(X_n, X_{n+k})/\text{var}[X_n]$ の漸近的性質 ($k \rightarrow \infty$) により、長記憶性が定義される。この X_n の自己相関係数は $\rho(k) = 1/2[(k+1)^{2H} - 2k^{2H} + (k-1)^{2H}]$ で与えられる。Leland [55, 56] や Willinger [98] は、この長記憶性を用いて自己相似性を次のように定義している。確率過程 $\{X_n; n = 1, 2, \dots\}$ があり、その平均を μ 、分散を σ^2 、自己相関係数を $\rho(k)$ とするとき、この確率過程 X_n は長記憶性をもつとする。そして、もとの確率過程の集積過程と呼ばれる過程 $\{X_i^{(m)}; i = 1, 2, \dots; m = 1, 2, \dots\}$ を考える。集積過程の分散が $\text{var}[X_i^{(m)}] = m^{-\beta}$ となり、かつ自己相関係数 $\rho^{(m)}(k)$ がすべての m に対して $\rho^{(m)}(k) = \rho(k)$ となるとき、この確率過程 X_n は厳密な 2 次の自己相似性をもつといい、十分大きな k に対して、 $\lim_{m \rightarrow \infty} \rho^{(m)}(k) = \rho(k)$ が成り立つとき、漸近的な 2 次の自己相似性をもつという。自己相関係数を用いる定義とは別の定義として、スペクトル密度関数 $S(f)$ を用いる方法もある。スペクトル密度関数はピリオドグラムより得られ、周波数成分 $f \rightarrow 0$ における漸近的性質として長記憶性が定義される。自己相関係数とスペクトル密度関数による定義は同値である [4]。

長記憶性をもつ理論的なモデルとして、Fractional Gaussian Noise (FGN) と Fractional Autoregressive Integrated Moving Average (fractional ARIMA) 過程が考えられる [4, 38, 57, 58]。FGN は FBM 過程の増分過程として定義される。FBM 過程は、通常の Brown 運動の加重平均として定義される [4]。また、FBM の別の定義として、自己相似パラメタが H である自己相似過程 $\{Y_n; n = 1, 2, \dots\}$ の増分過程 $\{X_n; n = 1, 2, \dots\}$ が Gauss 過程であ

るとしたときの $\{Y_n; n = 1, 2, \dots\}$ が FBM 過程であると定義することもでき、そしてこの X_n を FGN と定義する [72]. 一方, fractional ARIMA 過程は, ARIMA モデルの差分演算子の次数 d を実数に置き換えたものとして定義される [38]. このとき, 自己相似パラメタ H と d との間には $H = d + 1/2$ の関係がある. これらの過程は, 情報通信網上で長記憶性をもつモデルとしてよく用いられている [63].

本論文では, Moving Picture Experts Group 2 (MPEG2) と呼ばれる動画像圧縮方式によって VBR 符号化されたデータを用いる. MPEG は国際標準の動画像圧縮方式である. MPEG1 は, 1990 年に勧告された ISDN 上の動画像伝送を可能にした H.261¹ と, 1993 年に国際標準化された静止画像に対する国際標準の Joint Photographic coding Experts Group (JPEG) の技術を取り入れ, 1992 年 11 月に標準化された. MPEG1 は, 主に蓄積メディアである Compact Disk-Read Only Memory (CD-ROM) で用いられる. MPEG1 を High Definition Television (HDTV) や VBR 符号化に対応させることにより, MPEG2 が 1994 年 11 月に標準化された. MPEG2 は主に放送に使われることを目的とし, 現在 SKY Perfect TV! 等の Communication Satellite (CS) 放送や Digital Video Disk (DVD) 等で既に利用され始めている. 更に, 低ビット率での符号化や Object 符号化方式を取り入れた² MPEG4 が標準化され, 1998 年 5 月に最終委員会草案が完成している. 現在では, コンテンツの検索を可能にすることができる MPEG7 の標準化が進められている.

本論文の第 3 章では, 実際の画像データをもとに, その画像データが自己相似性をもつかどうかを検証した. 自己相似性を解析する方法は, 多数研究されている. 主な解析方法は, R/S 解析 [3, 4, 40, 90], 集積過程の分散を用いる分散解析 [3, 4, 90], スペクトル解析 [3, 4, 71, 90], そして Whittle の最尤推定法 [4, 97, 101] 等である. このうち, 本論文では, R/S 解析, 分散解析, 及びスペクトル解析を用いて, 自己相似性の解析を行った.

Hurst [40] は, ナイル川の水位の変化のデータから, それが長記憶性をもつことを発見した. R/S 解析は, 彼によって提唱された長記憶性を示すパラメタ H を推定する方法である. Hurst はまず, $\{X_j; j = 1, 2, \dots\}$ を時刻 j にダムに入ってくる水の量とし, それを m 個ずつに分けたときの平均を $X_i^{(m)} = (1/m) \sum_{j=(i-1)m+1}^{(i-1)m+n} X_j$ とし, $\sum_{j=(i-1)m+1}^{(i-1)m+n} X_j - nX_i^{(m)}$ の最大値と最小値の差として adjusted range を定義した. これを標本標準偏差で割ったものを R/S 統計量とした. そして, m 値に対する R/S 統計量の値を両対数グラフにプロットしたとき, $\log R/S \approx H \log m + \text{定数}$ となることを発見した. この H のことを Hurst パラメタといい, このことから長記憶性は Hurst 効果と呼ばれるようになった. Hurst パラメタは前に出てきた自己相似パラメタ H と同じ値をもつ. この長記憶性の統計的性質に関する研究もかなり行われている [7, 21, 88].

分散解析は, 長記憶性のもつ漸近的性質を利用するものであり, R/S 解析と同様に頻りに用いられている. この方法は, 集積過程の分散 $\text{var}[X^{(m)}]$ を計算し, m に対するこの値を両対数グラフにプロットし, その傾きを見ることにより Hurst パラメタを推定する方法である.

¹現在でもテレビ会議システム等で用いられている.

²Object 符号化方式では, 符号化伝送すべき 3次元の空間情報を, 空間内に位置する人や建物等の各オブジェクトの加工, 編集を可能にしている [61].

スペクトル解析は、Fourier 変換を用いて得られたスペクトル密度関数から長記憶性を検証する方法であるが、 R/S 解析や分散解析のように Hurst パラメタ H の値を明示的に推定することはできない。しかし、 R/S 解析や分散解析では調べることができない非定常性について検証することができるという利点がある。通常、スペクトル解析は周波数成分に対するピリオドグラムを両対数グラフにプロットするが、本論文では平滑化を行った [37, 44]。平滑化とは、ピリオドグラムを周波数分割し、その区間で平均を取ることである。

今まで見てきた 3 つ方法では、Hurst パラメタの信頼区間を求めることができない。そこで、使われているのが Whittle の最尤推定法である [97]。この方法を用いることにより、Hurst パラメタ H の信頼区間を求めることができる。Whittle の最尤推定法では、次のようにして Hurst パラメタを推定する。 $f(x; \theta) = \sigma_\epsilon^2 f(x; (1, \eta))$ を Gauss 過程 X_n のスペクトル密度関数とする。ここで、 $\theta = (\sigma_\epsilon^2, \eta) = (\sigma_\epsilon^2, H, \theta_3, \dots, \theta_k)$ と書け、 $\theta_3, \dots, \theta_k$ は、過程の短記憶性をモデル化しているパラメタである。 σ_ϵ^2 は白色雑音の分散である。このとき、 $\int_{-\pi}^{\pi} \log(f(x; (1, \eta))) dx = 0$ が成り立つ。 $I(x) = (2\pi n)^{-1} |\sum_{j=1}^n X_j e^{ijx}|^2$ が与えられたとき、 $Q(\eta) = \int_{-\pi}^{\pi} [I(x)/f(x; (1, \eta))] dx$ を最小にする $\hat{\eta}$ を推定する。更に、 $\hat{\sigma}_\epsilon^2 = \int_{-\pi}^{\pi} [I(x)/f(x; (1, \hat{\eta}))] dx$ が与えられる。このようにして求められた $\hat{\theta} = (\hat{\sigma}_\epsilon^2, \hat{\eta})$ に対して、 X_n のスペクトル密度関数 $S(f)$ が $S(f) \approx c_s |f|^{-\alpha}$ (c_s : 定数)³ で与えられるならば、 $n^{1/2}(\hat{\theta} - \theta)$ は漸近的に正規分布に従う [4, 98, 101]。これにより自己相似パラメタ H の信頼区間を得ることができる。

Whittle の最尤推定法以外にも、ピリオドグラムを利用する方法については、文献 [71] に書かれている。また、最近ではウェーブレットを用いて Hurst パラメタを推定する方法が研究されている [1, 9, 41]。

1.2.3 自己相似過程の生成法

自己相似性をもつ過程を人工的に生成することができれば、シミュレーションや実際の網の性能評価を行う際に非常に便利である。その代表として考えられるのが、FGN と fractional ARIMA モデルである [57, 38]。このような方法以外にも、On/Off ソースを多重化することによるもの [89] や $M/G/\infty$ 待ち行列システムを用いるもの [55, 68, 98] がある。また、 $M/G/\infty$ 待ち行列システムを用いた画像トラヒックのモデル化に関する研究も報告されている [47]。 $M/G/\infty$ 待ち行列システムのサービス時間分布が Pareto 分布等の裾の長い分布であるとき、系内人数の過程は長記憶性をもつ [21, 51]。長い裾のサービス時間分布をもつ $M/G/\infty$ システムを大偏差の理論によって解析したのが [22, 66] である。本論文の第 4 章では、この $M/G/\infty$ 待ち行列システムの系内人数を用いた自己相似過程の生成法の有効性を検証した。

³第 2.3 節で厳密な定義を与えるが、 c_s は定数でなくてもよい。

1.3 待ち時間の高次モーメントに関する研究

従来のトラヒックをモデル化する際には、待ち行列モデルがよく用いられていた。待ち行列モデルで使われる Markov 性は、最近の情報通信網をモデル化には用いることができないが、解析的に非常に扱いやすい。本節では、待ち行列モデルにおける客の待ち時間の高次モーメントについて説明する。第 1.3.1 節では、ランダム順サービス規律によってサービスされる M/G/1 待ち行列の高次モーメントについて説明し、第 1.3.2 節では、対称なポーリングシステムの 2 次モーメントについて説明する。

1.3.1 M/G/1 ランダム順サービスの待ち時間の高次モーメント

M/G/1 待ち行列システムにおいて平均待ち時間はサービスの順番(先着順サービス, 後着順サービス, ランダム順サービス)には依存しない。しかしながら, 待ち時間の分布, 及び 2 次以上の高次モーメントはサービスの順番に依存する。先着順サービスと後着順サービスでは, これらの高次モーメントは比較的簡単に求めることができるが, ランダム順サービスの場合には, 非常に複雑な計算を含むために簡単に計算することはできない。このことが, 今までの文献において, 待ち時間の平均と 2 次モーメントまでしか計算されていない理由である。高次モーメントは歪度や尖度のように待ち時間の分布をより明確にする特徴づけるに役立つ。

本論文の第 5 章では, ランダム順サービスによる M/G/1 待ち行列システムの待ち時間の 3 次モーメントから 6 次モーメントまでを *Mathematica* を用いて明示的に求めた [83]。このプログラムは Conolly [19] による方法を用いる。Conolly の方法は Kingman [42] によって導かれた待ち時間分布の分布関数の Laplace-Stieltjes Transform (LST) を用いて求める方法である。Kingman が求めた分布関数の LST は文献 [18, 77, 81] にも書かれている。その方法とは別に, [13, 14, 20] では LST を用いない方法が示されている。Fuhrmann [26] は先着順サービスシステムと後着順サービスシステムの待ち時間のモーメントについて興味深い関係を示している。しかし, Fuhrmann の結果は 2 次モーメントまでの関係にとどまっていた。本研究では, ランダム順サービスの高次モーメントと Takagi ら [82, 84] によって既に求められた先着順サービス, 及び後着順サービスの高次モーメントを比較する [83, 85]。

1.3.2 ポーリングシステムの待ち時間の 2 次モーメント

ポーリングシステムとは, N 個の待ち行列が 1 つのサーバによって周期的にサービスされるシステムである。そのようなシステムは, 1 つの資源を順番に割り当てることによって多くのユーザーが資源を共有できるようになるので通信システムや生産システムの性能評価モデルとして応用することができる。サーバが待ち行列に訪れる度にサービスされる客に関して 2 つのサービス規律を考える。全処理式サービスシステムとゲート式サービスシステムである。この 2 つのサービス規律の違いは, サービス中に到着した客が現在サービスしてい

るサーバによってサービスされるか、次にサーバが巡回してきたときにサービスされるかの違いである。前者を全処理式サービスシステムといい、後者をゲート式サービスシステムという。通常の待ち行列システムにおいて、客の立場から最も関心があるのは待ち時間である。 $N = 1$ の特別な場合はサーバの遊休時間 (バケーション) をもつ M/G/1 待ち行列システムである。また、 $N = \infty$ のとき (各待ち行列の負荷は無限小とする)、このシステムを連続ポーリングモデルという。ポーリングシステムにおける初期の解析やアプリケーションの研究については [79, 80] で報告されている。

各待ち行列において、客が Poisson 過程に従って到着し、一般サービス時間分布に従ってサービスされ (M/G/1 待ち行列システムのように)、サーバの移動時間も一般分布に従う全処理式もしくはゲート式サービス規律をもつポーリングシステムを考える。このとき、各待ち行列の客の待ち時間の任意の次数のモーメントは、連立線形方程式の解として計算できることが示されている。しかしながら、待ち行列の数が 2 個で対称なシステムのようにいくつかの特殊な場合を除いて、平均待ち時間ですら明示的な形で求められていない。故に、多くの研究者は、平均待ち時間を求めるために計算アルゴリズムを効率良くする方法 [45]、線形方程式の数を減らす方法 [24, 16, 73]、疑似保存則と呼ばれる線形方程式の間の線形従属性を利用する方法 [11] 等を研究した。

本論文の第 6 章では、全処理式サービスとゲート式サービスで処理されるポーリングシステムに対し、待ち時間の 2 次モーメントを明示的に計算した [52]。2 次モーメントは、一般にばらつき具合を表し、確率分布を推定する Chebyshev の不等式にも使われる。本研究での 2 次モーメントの計算は、すべての待ち行列が統計的に同一であるような対称なシステムに制限している。 N 個の待ち行列をもつ対称なポーリングシステムの平均待ち時間は既に明示的に与えられている [33]。過去の研究において、待ち時間の 2 次モーメント (分散) はバケーションをもつサーバ ($N = 1$) の場合も含め、次のような特別な場合にしか求められていない: サーバの移動時間が 0 で待ち行列の数が 2 個の非対称な全処理式サービスシステム [78]、同様のシステムでサーバの移動時間が 0 でないシステム [75]、待ち行列の数が 2 個並びに 3 個の場合で、サービス時間と移動時間が一定のシステム [2]、対称な連続ポーリングシステム [17, 25]。分散を数値的に計算する方法は [25, 45, 73] に書かれている。

1.4 本論文の構成

本論文の第 2 章以降は次のように構成される。

まず第 2 章において、自己相似過程と長記憶性について説明をする。第 2 章は、第 3 章と第 4 章における情報通信網のトラヒックの自己相似性に関する解析を行う際に必要となる理論の説明である。

第 3 章の前半部 (第 3.1 節 – 第 3.3 節) は圧縮動画方式の MPEG についての説明である。第 2 章と第 3 章のこの部分までが既存の概念である。第 3 章の後半部 (第 3.4 節 – 第 3.6 節) は測定データの解析方法とその解析結果を示す。この章で、MPEG2 によって圧縮さ

れた動画像は長記憶性をもっていることを確かめることができた。

第4章では、 $M/G/\infty$ 待ち行列システムを使った自己相似過程の生成法を説明し、そのようにして生成された自己相似過程の妥当性をシミュレーションによって検証する。第4章までが、情報通信網を流れるトラフィックの解析である。

第5章と第6章では、Markov 過程に基づく待ち行列モデルの解析を行う。これらのモデルは情報通信網のトラフィックが自己相似性と呼ばれる性質をもつといわれ始める前にこれらの網をモデル化する際に用いられていたものである。最近の情報通信網を流れるトラフィックの性能評価等の解析では、このような Markov 性を仮定した待ち行列モデルを用いたモデル化はできないが、それらは解析的に扱いやすい。実際 ボーリングシステムは FDDI や トークンリングの性能評価をする際に用いられている。

第5章では、ランダム順サービスをもつ $M/G/1$ 待ち行列の待ち時間分布の高次モーメントの解析を行い、2乗変動係数、歪度、及び尖度の観点から先着順サービスと後着順サービスとの比較を行う。

第6章は対称なボーリングシステムの待ち時間分布の2次モーメントの解析を行う。待ち行列の数が任意の場合についての結果は得られなかったが、ボーリングシステムの待ち行列の数 N が無限である連続ボーリングシステムと比較することにより、待ち行列の数が一般の N の場合の上限と下限を与える。

最後に、第7章において、今後の課題を示唆する。

1.5 研究の概要

本研究は、Asynchronous Transfer Mode (ATM: 非同期転送モード) 網等の高速情報通信網を流れるトラフィックの特性について研究したものである。昨今、インターネットや SKY Perfect TV! 等の通信衛星を利用した画像配信サービス (Communication Satellite (CS) 放送) において、動画像をはじめ多種多様なデータが網上を流れている。そのようなデータは従来の電話やパケット交換とは異なり、自己相似性と呼ばれる特性を持つといわれている。

本論文では、まず、この自己相似性と呼ばれる性質について説明をする。自己相似性とは、ある時系列データ $\{Y_n; n = 1, 2, \dots\}$ の時間単位を変えたとき、 Y_{cn} の分布がもとの時系列データの分布の形と似ている性質である。このような時系列データは、その増分過程 ($X_n = Y_n - Y_{n-1}$) が長記憶性と呼ばれる性質を持つ。長記憶性とは、ある過程 $\{X_n; n = 1, 2, \dots\}$ の自己相関係数 $\rho(k) = \text{cov}(X_n, X_{n+k}) / \text{var}[X_n]$ が、 k に対して非常にゆっくりと減少していく性質である。このことは、スペクトル密度関数 $S(f)$ を用いることによっても説明することができる。スペクトル密度関数を用いた場合、原点近傍の漸近的な振る舞いが $f^{-\alpha}$ ($0 < \alpha < 1$) に比例するということによって説明できる。実際の時系列データからこのような性質が検出されれば、自己相似性を有するということができる。

そこで、実際に Moving Picture Experts Group 2 (MPEG2) 方式によって圧縮されたデータを用いて、この性質を検証した。MPEG2 とは、動画像を圧縮させる方式であり、Digital

Video Disk (DVD) のような蓄積メディアをはじめ, 上述の CS 放送にも使用されている. 更に最近では, パソコンレベルで符号化復号化できる周辺機器も発売されている.

本論文では, MPEG2 の符号化形式について簡単に説明し, 続いて, 長記憶性を調べる手法について説明する. 本論文で用いた解析方法は Hurst によって提案された R/S 解析, 集積過程を用いた分散解析, そしてスペクトル密度関数を利用したスペクトル解析である. 解析に用いたデータは資料映像の Laser Disk (LD) 3 本と, そのうちの 1 本をソフト符号化したもの, そして 10 時間の地上波放送画像である. R/S 解析と分散解析は, 3 本の LD を順番に繋げたもの (直列データ), 多重化を想定して同じ時刻のものを足し合わせたもの (並列データ), 2 つを順番に繋げその上に乱数でずらして別のもう 1 本を足し合わせたもの (混合データ) のようにして, LD の画像のみに対して行った. スペクトル解析は, 上の 3 つデータとは別に一つ一つの LD データと放送画像について行った. それぞれのデータについてフレーム単位 (1/30 秒間隔) と Group of Pictures (GOP) 単位 (1/2 秒間隔) で行った. その結果, 自己相似性を確認することができた.

しかしながら, このように符号化されたデータを扱うことができるのはかなり希である. そこで, 簡単な待ち行列システムを用いることにより自己相似性をもつ時系列データを人工的に作る方法を, シミュレーションにより検証した. この方法は $M/G/\infty$ 待ち行列システムを用いる方法である. まず $M/G/\infty$ 待ち行列システムの時刻 t と時刻 $t+l$ における系内人数の相関係数を計算する. 時刻 0 にシステム内の系内人数は Poisson 分布に従い, それらの客の残りのサービス時間分布はもとのサービス時間分布の余命であるとする. サービス時間分布としては Pareto 分布を用いる. 系内人数の相関係数を用いることによって, 与えられた自己相似パラメタをもつ時系列データを生成することができる. シミュレーションの結果, $H \lesssim 0.7$ の範囲内にあるとき, この方法を使えることが分かった.

これらの情報通信網のトラヒックの特性に関する一連の研究の他に, 基本的な待ち行列における待ち時間の分布の高次モーメントに関する研究も行った. 1 つはランダム順サービスをもつ $M/G/1$ 待ち行列システムにおける待ち時間の高次モーメントの計算である. 先着順サービスや後着順サービスをもつ $M/G/1$ 待ち行列システムの待ち時間の高次モーメントの計算は比較的簡単に行うことができるが, ランダム順サービスの場合, その Laplace-Stieltjes Transform (LST) が複雑な形をしているため, 2 次モーメントまでしか与えられていない. そこで数式処理ソフト *Mathematica* を用いてこの高次モーメントを計算した. 更に, その結果から先着順サービスと後着順サービスの 2 乗変動係数, 歪度, 尖度との比較を行った.

もう 1 つは, 対称なポーリングシステムにおける待ち時間の 2 次モーメントの計算である. ポーリングシステムとは, 1 つのサーバが巡回して N 個の待ち行列をサービスするシステムである. この研究では, 対称なポーリングシステムの待ち時間の 2 次モーメントを, 全処理式サービス方式とゲート式サービス方式について, 待ち行列の数 N が $N = 2, 3, 4$ の場合について計算した. 全処理式サービスとは, サーバが巡回したときに待ち行列が空になるまでその待ち行列をサービスし, 待ち行列が空になったら次の待ち行列に移動するサービス方式である. ゲート式サービスとは, サーバが巡回してきたときにその待ち行列にいる客だ

けをサービスする方式である。サーバがサービス中に到着した客は次にサーバが回ってくるまでサービスされない。一般の N については方程式が複雑なため解くことができなかった。しかしながら、待ち行列の数 N が無限大である連続型ポーリングシステムと $N = 1$ のポーリングシステムを比較することにより、一般に N 個の待ち行列をもつポーリングシステムの待ち時間の 2 次モーメントについて、かなり厳しい上界と下界を与えることができた。