

情報ネットワークにおける通信リクエスト発生の非定常ポアソン性検定

篠 秀明 北澤 慶一 治部 将之 原 淳平 八名 和夫
法政大学工学部電子情報学科

あらし

WWW(World Wide Web)の利用を中心としたインターネットユーザが急増する現在において、通信トラヒックの特性を把握し、必要なネットワーク資源を推定することはネットワーク構築・運用の立場から重要な課題である。ネットワークトラヒック時系列の統計的なモデル化がなされればシステム全体の負荷変動特性等ネットワークの振る舞いをあらかじめ検討することが可能となり上記課題の解決に有効な手段を提供することになると思われる。WWWトラヒックモデル化の基本的な要素である通信リクエストの発生時刻列の特徴付けについて、いくつかの報告があるが、過程の非定常性を考慮に入れた詳細な検討はなされていない。本論文ではネットワークにおける通信リクエスト発生の統計的性質について検討し、特に非定常過程としての特徴付けの方法とその解析例を、シミュレーションデータ、実データを用いて示した。

1 まえがき

近年のWWWの利用によるインターネットへの接続端末数の増加に伴い、ネットワーク上には大量のトラヒックが生じている。データの転送に用いられる通信回線等のネットワーク資源は限られたものであり、その有効利用のためにネットワークトラヒックを定量的に評価することはネットワークの設計・評価において重要な課題である。ネットワークトラヒックの詳細な統計的モデル化が行われれば経験に依存することの多い実際のネットワーク構築以前にシステム全体にわたるトラヒックの評価等が可能となり上記課題の解決に有効な手段を提供することとなろう。本論文ではネットワークトラヒックモデル化において重要な要素となるトラヒックリクエストの発生時刻系列について、ProxyサーバからWWWトラヒックのリクエスト発生時刻系列を取り出しその統計的性質の解析を行った。従来WWWトラヒックにおけるリクエストの生起は対数正規分布に従うと言う報告[1]がなされているが、トラヒック時系列の非定常性を考慮した詳細な検討はなされていない。本報告ではリクエスト発生過程の非定常性を考慮した特徴づけについて論じている。特に、時間尺度変換を用いた非定常点過程に対するポアソン性検定法をシミュレーションデータ及び実際のトラヒックデータに適用し、トラヒックリクエスト発生時刻列が非定常ポアソン過程としてモデル化可能であるかどうかを詳細に検討した。

2 非定常過程のポアソン性検定

定常な点過程において、ポアソン過程は過去の点過程の実現が将来の事象生起に影響を与えないという性質がある。しかし、非定常な場合では誤って定常性を仮定して解析を行なうと非ポアソン過程であるとの誤った結論に至る可能性がある。そこで、過程が非定常ポアソン過程であるとの

帰無仮説のもとで事象生起密度関数 $\lambda(t)$ に対し、

$$\Lambda(t) \equiv \int_0^t \lambda(u) du \quad (1)$$

なる時間尺度変換を行ない定常点過程に対するポアソン性検定に帰着する方法を用いる。パラメータ関数が $\Lambda(t)$ であるような非定常ポアソン過程の1つの実現を $\{t_1, \dots, t_N\}$ とするとき

$$u_k \equiv \Lambda(t_k) \quad ; k = 1, \dots, N \quad (2)$$

とすれば $\{u_k\}_{k=1}^N$ は平均事象生起間隔1の定常ポアソン過程の1つの実現とみなすことができ、逆に平均事象生起間隔1の定常ポアソン過程の1つの実現 $\{u_1, \dots, u_N\}$ に対し、

$$t_k \equiv \Lambda^{-1}(u_k) \quad ; k = 1, \dots, N \quad (3)$$

と定義すれば $\{t_k\}_{k=1}^N$ はパラメータ関数を $\Lambda(t)$ とする非定常ポアソン過程の1つの実現とみなせることが知られている。

ここで、このような時間尺度変換を導入すると非ポアソン過程に対してもうまく変換のための関数を選べばポアソン過程に変換されてしまうのではないかという疑問が生ずる。点過程全体として過程の過去の実現(以下単に履歴という) $\mu(t)$ を条件とする事象生起密度 $\lambda(t|\mu(t))$

$$\lambda(t|\mu(t)) \equiv \lim_{h \downarrow 0} \frac{\Pr[\Delta N_t = 1 | \mu(t)]}{\Delta t} \quad (4)$$

が存在し条件付き希少性

$$\lim_{h \downarrow 0} \frac{\Pr[\Delta N_t \geq 2 | \mu(t)]}{\Pr[\Delta N_t = 1 | \mu(t)]} = 0 \quad (5)$$

を満たすものとする。ここで、 N_t は時刻 t までの計数過程であり、また

$$\Delta N_t \equiv N_{t+\Delta t} - N_t \quad (6)$$

である。しかし文献 [2] において履歴 $\mu(t)$ が依存しない場合、すなわち

$$\lambda(t) = \lim_{h \downarrow 0} \frac{\Pr[\Delta N_t = 1 | \mu(t)]}{\Delta t} \quad (7)$$

における場合においては、非ポアソン性はいかなる時間尺度変換に対しても保存されることが明らかになっている。

以上より、過去の履歴に直接依存しない事象生起密度関数 $\lambda(t)$ に時間尺度変換を行ない、実現された見本過程に対し各種の定常ポアソン性の検定を行なうことにより非定常過程を考慮したポアソン性検定が構成できる。

3 事象生起密度関数のパラメータ推定

3.1 非定常 Poisson 過程の尤度関数

事象生起密度のパラメトリックな強度関数が与えられ、観測される点過程の 1 つの見本過程に基づくパラメータ推定を行おうとする場合において尤度関数が必要となる。

観測時間 $[0, T]$ における事象生起数を $N(T)$ 、生起時刻を $T_1, T_2, \dots, T_{N(T)}$ とする。また、それらの実現値を n, t_1, t_2, \dots, t_n とする。このとき、 T_1, T_2, \dots, T_n の結合確率密度を $f(t_1, t_2, \dots, t_n; N(T) = n)$ とすれば、

$$\begin{aligned} & f(t_1, t_2, \dots, t_n; N(T) = n) dt_1, \dots, dt_n \\ &= e^{-\int_0^{t_1} \lambda(\xi; \theta) d\xi} \lambda(t_1) dt_1 \cdot e^{-\int_{t_1}^{t_2} \lambda(\xi; \theta) d\xi} \lambda(t_2) dt_2 \cdot \\ & \quad \dots \cdot e^{-\int_{t_{n-1}}^T \lambda(\xi; \theta) d\xi} \lambda(t_n) dt_n \\ &= \prod_{i=1}^n \lambda(t_i) e^{-\int_0^T \lambda(\xi; \theta) d\xi} dt_1 \dots dt_n \quad (8) \end{aligned}$$

で与えられる。ここで $\lambda(\xi; \theta)$ はパラメトリックな強度関数を表す。式 (8) に事象生起時刻を代入して θ の関数と見たものが尤度関数である。尤度関数の自然対数をとった対数尤度関数を $\mathcal{L}(\theta)$ とすると

$$\mathcal{L}(\theta) = -\int_0^T \lambda(\xi; \theta) d\xi + \sum_{i=1}^n \ln \lambda(t_i; \theta) \quad (9)$$

となる。

3.2 最尤推定

定常ポアソン過程における生起事象間隔は指数分布に従うので、事象生起密度関数が

$$\lambda(\xi; \theta) = \theta_1 \cdot e^{\theta_2 \xi} \quad ; \theta_1 > 0 \quad (10)$$

であるようなモデルを立てる。式 (9) より対数尤度関数 $\mathcal{L}(\theta)$ は

$$\begin{aligned} \mathcal{L}(\theta) &= -\int_0^T \theta_1 e^{\theta_2 \xi} d\xi + \sum_{i=1}^n (\ln \theta_1 + \theta_2 t_i) \\ &= \frac{\theta_1}{\theta_2} (1 - e^{\theta_2 T}) + n \ln \theta_1 + \theta_2 \sum_{i=1}^n t_i \quad (11) \end{aligned}$$

となる。ここで上式 (11) を最大とするパラメータは最適なモデルを与える。このような方法を最尤推定と呼ぶ。パラメータ θ_1, θ_2 の最尤推定量 $\hat{\theta}_1, \hat{\theta}_2$ は尤度方程式より、 $\partial/\partial\theta_1 = 0, \partial/\partial\theta_2 = 0$ とした次式

$$\sum_{i=1}^n t_i + n \left(\frac{1}{\theta_2} - \frac{T e^{\theta_2 T}}{e^{\theta_2 T} - 1} \right) = 0 \quad (12)$$

$$\theta_1 = \theta_2 \frac{n}{e^{\theta_2 T} - 1} \quad (13)$$

を連立させて解くことにより得られ、これがモデル式 (10) に対する最尤パラメータとなる。

4 シミュレーション

実データの解析を行なう前に、シミュレーションにより事象生起密度関数のモデル (式 (10)) のもとで最尤推定によりパラメータが良好に推定されていることを示す。さらに定常ポアソン性の検定方法の幾つかを挙げ、シミュレーションデータに対し検定を行なうことにより、最尤パラメータから得られるパラメータ関数 (式 (1)) による時間尺度変換を施した点生起時刻列が、定常単位生起密度をもつ一様ポアソン過程からの見本過程と見なせることを示す。

4.1 データ作成方法

以下にシミュレーションデータ作成手法を示す。

1. 定常単位ポアソン過程からの見本過程を作成するには、点生起時刻列の間隔は指数分布に従うので、指数分布に従う乱数を作成する。指数分布に従う乱数は、その確率分布関数

$$F(x) = \int_0^x e^{-x} dx \quad (14)$$

の逆関数

$$F^{-1}(x) = -\ln(1 - x) \quad (15)$$

で一様乱数 $U (0 \leq U < 1)$ を変換して $-\ln(1 - U)$ として生成する。こうして作成されたデータ系列を w とする。

2. 時刻列 w に対して

$$\lambda(t; \theta) = \theta_1 \cdot e^{\theta_2 t} \quad ; \theta_1 > 0 \quad (16)$$

なる事象生起密度をもつ非定常ポアソン過程からなる見本過程を作成する。

4.2 結果

ランダムに設定したパラメータ

- $\theta_1 = 100, \theta_2 = 20$ (data1)
- $\theta_1 = 10, \theta_2 = 1$ (data2)
- $\theta_1 = 5, \theta_2 = 5$ (data3)

をもつデータ点数 1000 点の事象生起密度関数時刻列の見本過程を 4.2 節の方法でそれぞれ作成し、シミュレーションに用いた。

4.2.1 最尤推定

最尤推定により求められた最尤パラメータ $\hat{\theta}_1, \hat{\theta}_2$ を表 1 に示す。

表 1. 最尤推定により求められた最尤パラメータ

	パラメータ (θ_1, θ_2)	最尤パラメータ $(\hat{\theta}_1, \hat{\theta}_2)$
data1	(100, 20)	(102.508, 19.941)
data2	(10, 1)	(10.239, 0.997)
data3	(5, 5)	(5.121, 4.990)

推定された最尤パラメータ $\hat{\theta}_1, \hat{\theta}_2$ は、データ作成のために設定したパラメータに非常に近い値となり、式 (10) のモデルに対する事象生起密度関数の最尤推定がよく行なわれていることがいえる。よって、時間尺度変換 (式 (1)) を用いて求められる見本過程に種々の定常性ポアソン性検定を行なうことにより、非定常過程を考慮したポアソン性検定として見なせることが示された。

4.2.2 計数過程

観測する過程が始まる任意の時間原点からの累積事象数を、各シミュレーションデータごとに左側に時間尺度変換前、右側に時間尺度変換後として図 1~図 3 に示す。

観測時間原点からの累積事象数である計数過程 N_t は、式 (7) で表せるように、事象生起密度を 2 次の統計量とみなしたものになる。時間尺度変換前ではデータ作成に用いた指数の傾向変化が反映されているが、時間尺度変換後の見本過程は事象生起密度が一様に分布していることがわかる。

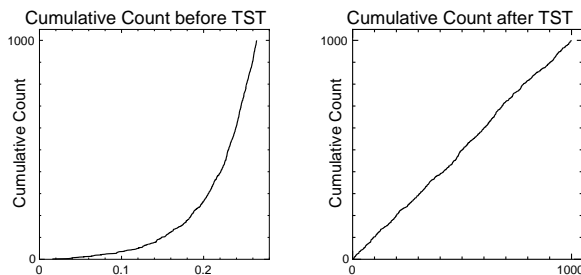


図 1. $\theta_1 = 100, \theta_2 = 20$ 累積カウント数 (左は時間尺度変換前、右は時間尺度変換後)

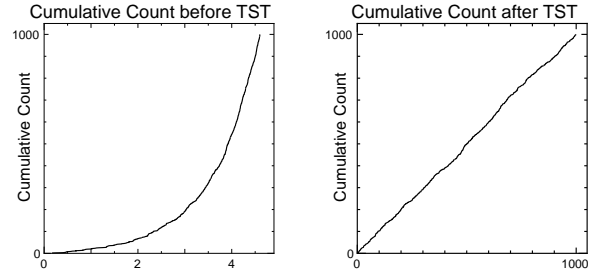


図 2. $\theta_1 = 10, \theta_2 = 1$ 累積カウント数 (左は時間尺度変換前、右は時間尺度変換後)

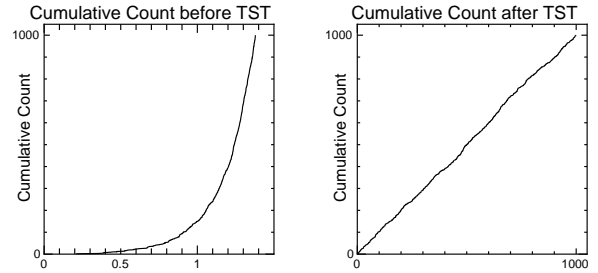


図 3. $\theta_1 = 5, \theta_2 = 5$ 累積カウント数 (左は時間尺度変換前、右は時間尺度変換後)

4.2.3 インターバル分布

再生過程理論の立場から、ポアソン過程は指数分布を共通の点間隔密度としてもつ再生過程といえる。点間隔の標本分布と指数分布の適合性を χ^2 検定等によりポアソン性検定が行なえる。表 2 に χ^2 分布との適合度を、インターバル分布と指数分布の適合性を視察するため図 4~図 6 に分布図を示す。なお、ヒストグラムのビン数は 30、ビンの幅は平均点間隔の 5 分の 1 である。各シミュレーションデータごとに左側は時間尺度変換前、右側は時間尺度変換後であり、図中の実線は観測されたインターバル分布、破線は指数分布を表す。

すべてのシミュレーションデータに対し、時間尺度変換前では各データとも有意水準を大きく越え棄却されたが、時間尺度変換後では有意水準内に収まり棄却されなかった。

表 2. χ^2 適合度検定

	時間尺度変換前	時間尺度変換後
data1	406.79	24.04
data2	344.29	22.97
data3	684.72	22.11
5% 有意水準	43.8	43.8

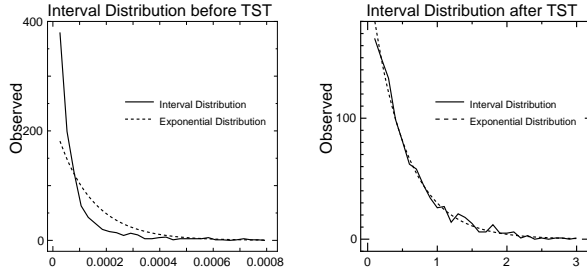


図 4. $\theta_1 = 100, \theta_2 = 20$ インターバル分布 (左は時間尺度変換前、右は時間尺度変換後)

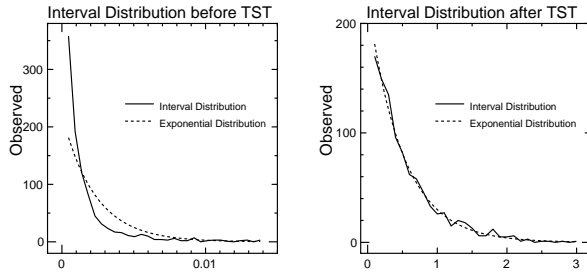


図 5. $\theta_1 = 10, \theta_2 = 1$ インターバル分布 (左は時間尺度変換前、右は時間尺度変換後)

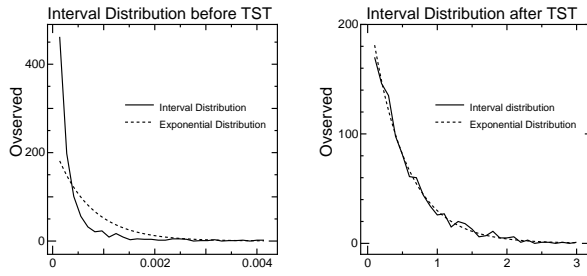


図 6. $\theta_1 = 5, \theta_2 = 5$ インターバル分布 (左は時間尺度変換前、右は時間尺度変換後)

4.2.4 系列相関係数

系列相関係数は以下の方法により推定される。間隔系列 $\{\tau_1, \tau_2, \dots, \tau_n, \dots\}$ が与えられたとき、系列共分散関数の推定値 \tilde{C}_k は

$$\tilde{C}_k = \frac{1}{n-k} \sum_{i=1}^{n-k} (\tau_i - \bar{\tau}')(\tau_{i+k} - \bar{\tau}'') \quad (17)$$

となる。ただし、

$$\bar{\tau}' = \frac{1}{n-k} \sum_{i=1}^{n-k} \tau_i \quad (18)$$

$$\bar{\tau}'' = \frac{1}{n-k} \sum_{i=1}^{n-k} \tau_{i+k} \quad (19)$$

である。系列相関係数 $\hat{\rho}_k$ は系列共分散関数の推定値 \tilde{C}_k を用いて以下のように推定量に推定される。

$$\hat{\rho}_k = \frac{\tilde{C}_k}{(\tilde{C}'_{0,k} \cdot \tilde{C}''_{0,k})^{1/2}} \quad (20)$$

ただし、

$$\tilde{C}'_{0,k} = \frac{1}{n-k} \sum_{i=1}^{n-k} (\tau_i - \bar{\tau}'_k)^2 \quad (21)$$

$$\tilde{C}''_{0,k} = \frac{1}{n-k} \sum_{i=1}^{n-k} (\tau_{i+k} - \bar{\tau}''_k)^2 \quad (22)$$

$\hat{\rho}_k$ は漸的に正規分布に従う不偏推定量であり分散は漸的に

$$var(\hat{\rho}_k) = \frac{1}{n-k} \sum_{i=-\infty}^{\infty} \rho_i^2 \quad (23)$$

となる。式 (23) より $\hat{\rho}_k$ を用いて実際に観測されるデータが、定常性の仮定のもとで再生過程とみなせるかどうかの検定が構成できる。帰無仮説を「見本過程 τ_1, \dots, τ_n が再生過程からのサンプルである」としたとき、任意の $k (< n)$ に対して

$$|\hat{\rho}_k| > \frac{C_{1/2\alpha}}{\sqrt{n-k}} \quad (24)$$

であれば、帰無仮説は棄却される。ここで、 $C_{1/2\alpha}$ は単位正規分布の上側 $1/2\alpha$ 点である。

各シミュレーションデータにおける推定された系列相関係数を、左側には時間尺度変換前、右側には時間尺度変換後と図 7～図 9 に示す。なお、ラグは 100 とし、内側の破線は 5%、外側の破線は 1% の有意水準を表す。

図 7～図 9 より時間尺度変換前における系列相関係数は長く続く正の相関があることが読みとれ、有意水準内に収まらず帰無仮説は棄却された。それに対し、時間尺度変換後においてシミュレーションではいずれも有意水準を越えたラグの数は 1% 有意水準内 (1%)、5% 有意水準内 (3%) となり帰無仮説は棄却されなかった。

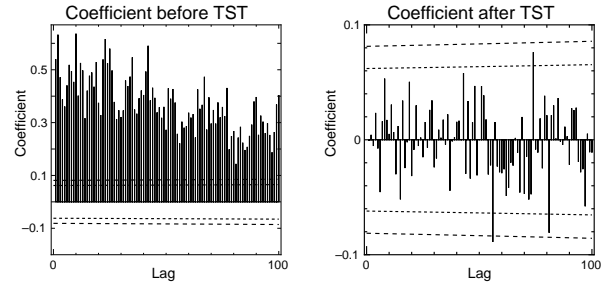


図 7. $\theta_1 = 100, \theta_2 = 20$ 系列相関係数 (左は時間尺度変換前、右は時間尺度変換後)

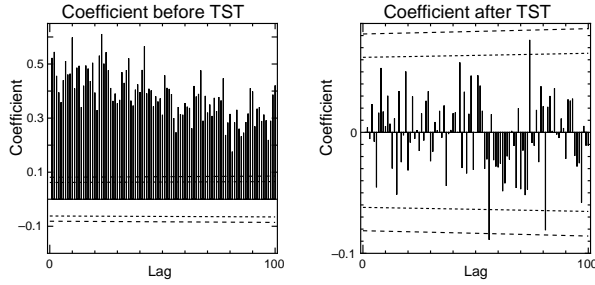


図 8. $\theta_1 = 10, \theta_2 = 1$ 系列相関係数 (左は時間尺度変換前、右は時間尺度変換後)

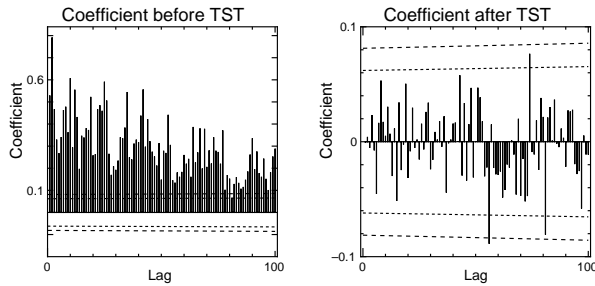


図 9. $\theta_1 = 5, \theta_2 = 5$ 系列相関係数 (左は時間尺度変換前、右は時間尺度変換後)

4.2.5 生起事象数に関する検定

時間軸を等間隔 Δ に分割し各区間に生じた事象数を $n_i; i = 1, \dots, K$, n_i の標本平均を \tilde{n} とする。 n_i の \tilde{n} からの拡がりを

$$d = \sum_{i=1}^k \frac{(n_i - \tilde{n})^2}{\tilde{n}} \quad (25)$$

とすると、帰無仮説のもとで d は自由度 $K - 1$ の χ^2 分布で近似されることから、1つのポアソン過程の検定が構成される。 d を $K - 1$ で除した量は事象生起個数の分散と平均事象生起数の比であり、ポアソン過程における理論値 1 との大雑把な比較にあたる。

各シミュレーションデータにおける時間尺度変換前後 d の値を表 3 に示し、時間尺度変換前後での比較を行なう。なお、瓶の数は 10 とし、表 3 中の () 内は $d/9$ の値である。

表 3. 生起事象数に関する検定

	時間尺度変換前	時間尺度変換後
data1	1620.07(180.01)	10.6997(1.189)
data2	1318.04(146.45)	10.6997(1.189)
data3	2329.60(258.84)	10.4000(1.156)

$d/9$ の値について、時間尺度変換前ではばらつきがある大きな値であったのに対し、時間尺度変換後でポアソン過程の理論値である 1 に近い値 (1 ± 0.2 の範囲) を得た。

4.2.6 Sherman 統計量

観測時間 T を 1 とし正規化した生起時刻を $v_i = t_i/T$ とすると、 $N_T = n$ の条件のもとで

$$E[v_i - v_{i-1}] = \frac{1}{n+1} \quad i = 1, \dots, n+1 \quad (26)$$

となる。ただし $t_0 = 0, t_{n+1} = T$ とする。 Sherman 統計量は $v_i - v_{i-1}$ の期待値からの偏差を累計し

$$S_n = \frac{1}{2} \sum_{i=1}^{n+1} \left| v_i - v_{i-1} - \frac{1}{n+1} \right| \quad (27)$$

と定義される。 Sherman 統計量はインターバルが一定に近くなるほど小さな値をとり、不規則さが増すほど大きな値をとる。ポアソン過程のもとでは

$$E[S_n] = \frac{1}{e} \quad (28)$$

$$\text{var}[S_n] = \frac{0.05908}{n} - \frac{0.07145}{n^2} + o\left(\frac{1}{n^3}\right) \quad (29)$$

となり正規化された Sherman 統計量は漸近的に正規分布となることが知られている。よって、式 (28)、式 (29) を用いてポアソン性の検定が行なえる。

各シミュレーションデータの時間尺度変換前後の Sherman 統計量を表 4 に示す。なお、各統計値下の値は式 (28)、(29) より求められる 95% 信頼区間である。

表 4 より、時間尺度変換前ではすべてのデータとも 95% の信頼区間内からはずれ帰無仮説は棄却された。それに対し時間尺度変換後においては 95% の信頼区間内に収まり、帰無仮説は棄却されなかった。

表 4. Sherman 統計量

	時間尺度変換前	時間尺度変換後
data1	0.5806	0.3670
95% 信頼区間	(0.3528, 0.3829)	(0.3528, 0.3829)
data2	0.5515	0.3669
95% 信頼区間	(0.3528, 0.3829)	(0.3528, 0.3829)
data3	0.6374	0.3670
95% 信頼区間	(0.3528, 0.3829)	(0.3528, 0.3829)

4.3 考察

最尤推定により得られた最尤パラメータ $\hat{\theta}_1, \hat{\theta}_2$ はデータ作成のために設定した値とほぼ一致し、良く推定されていることが示せた。時間尺度変換前においては系列相関係数は長く続く強い正の相関を持ち、また χ^2 適合度、 Sherman 統計量で帰無仮説は棄却された。これは過程の非定常性が及ぼした影響によるものである。それに対し、非定常性を考慮した時間尺度変換後の結果では $d/9$ はポアソン過程の理論値 1 に近い値をとり、系列相関係数、 χ^2 適合度、 Sherman 統計量とも有意水準内に収まり帰無仮説は棄却されなかった。すなわち、最尤パラメータから得られるパラメータ関数 (式 (1)) による時間尺度変換を施した点生起時刻列

が、定常単位生起密度をもつ一様ポアソン過程からの見本過程と見なせる。

なお、 $(\theta_1 = 1, \theta_2 = 0.1)$ 、 $(\theta_1 = 20, \theta_2 = 1)$ 、 $(\theta_1 = 50, \theta_2 = 5)$ の場合についても、最尤推定によりそれぞれ指定した値に非常に近いパラメータが求められ、それら事象生起密度をもつ時刻列に対し検定を行なったが、時間尺度変換後において本章で挙げたすべての有意性検定法で棄却されなかった結果が明らかになっている。

以上より、実際の点過程データの事象生起密度関数が式(10)のモデルに適合している場合において、非定常ポアソン過程を検定することが可能である。

5 実データ解析

今回の解析に用いたデータは、法政大学小金井キャンパスに設置されている Proxy サーバ、squid のアクセスログ中のユーザリクエストに直接関係する TCP/IP に関する情報を取り出したものである。このアクセスログ中にはリクエスト発生ユーザ、リクエスト先、どのように扱われて処理されたかを示すログタグ、データオブジェクトの大きさ、データの転送開始時間、終了時間等が記録されており、リクエストの生起時刻や、転送時間、処理方法と言った情報を得る事ができる。本章では、データ概要とそのデータ加工手法について述べ、得られた実際のリクエスト発生時刻列に非定常ポアソン性の検定を行なう。検定に際しては前述と同じく、モデルを式(10)とし最尤推定を行なう。これにより得たパラメータをもつ事象生起密度関数に時間尺度変換(式(1))を施し、定常ポアソン過程と見なせる非定常過程の見本過程を作成する。そこで、4.3節で述べた定常ポアソン性の検定法により非定常性のポアソン性の検定を行なう。

5.1 実データ概要

アクセスログの実データは以下の10のフィールドに分かれている。

その一例を以下に示す。

```
909500409.8701 7862 liberty.ant.k.hosei.ac.jp3  
TCP_MISS/2004 14535 GET6  
http://www.disneyblast.com/ActiveDesktop7  
-8 NONE/-9 application/x-netcdf10
```

各フィールドの概要は以下のように定められている。

1. タイムスタンプ
クライアントソケットが閉じられた時間。単位は秒で、UNIX 時間で記録される。
2. 経過時間
クライアントからの要求経過時間。単位はミリ秒。クライアントソケットの accept() から close() までの時間。
3. クライアントアドレス
接続したクライアントの IP アドレス。squid.conf 中の log_fqdn 項目が有効になっていれば、FQDN で

記録される。

4. ログタグ / HTTP コード番号
ログタグは、どこでどのように扱われ処理されたのかを示している。HTTP コード番号は、HTTP 応答ヘッダの最初の行から取られる。ICP の要求の場合は 000 となる。また、squid.conf の log_icp_queries が無効になっている場合には、ICP の要求については記録されない。
5. 大きさ
データオブジェクトの大きさ (バイト数)。
6. 要求方法
GET、PUT 等の HTTP の要求方法。ICP 要求の場合は、ICP_QUERY となる。
7. URL
要求された URL。
8. Indent
squid.conf 中の ident_lookup が有効ならば、このフィールドが活用され、Ident サービスから得られた情報が記録される。無効ならば、"- "が記録される。
9. 階層構造データタグ / ホスト名
どのようにオブジェクトが取得されたか、そして、そのオブジェクトがどこから取得されたのかを記録する。
10. コンテンツタイプ
HTTP 応答ヘッダの Content-Type フィールドから取得する。

今回の実データ解析に用いるデータは、「ユーザのリクエストに直接関わる TCP 要求のみのリクエスト発生時刻列」であるため、注目すべき項目は1、2、4の3つの項目である。特に注意すべき点としては、転送終了時間と経過時間の単位が [sec] と [msec] であることと、ログタグの処理方法情報が TCP のもののみを抽出する必要があるということである。

5.2 実データ加工手法

次に、解析を行なうためのアクセスログデータの加工手段について述べる。

1. アクセスログ中の4. ログタグが UDP のものを除去する。
2. リクエスト発生時間を以下の式から求める。
リクエスト発生時間 [sec] =
伝送終了時刻 (タイムスタンプ:UNIX time)[sec] -
0.001 * 転送 (経過) 時間 [msec]
3. リクエストの発生を時刻順にソートする。
4. 解析能単位が [msec] なので、同時刻に生起しないよう [msec] 以下の部分に - 0.5[msec] ~ 0.5[msec] の一様乱数を付加する。
5. リクエストは性質上発生がバースト的になるため同一ユーザの最初のリクエストから 2[sec] 以内のものを無視し、1つのリクエストとしてグループ化する。

以上の手法により、リクエストの時刻列データファイルが作成される。このようにして得られたリクエスト生起時刻列に対し、非定常性を考慮したポアソン性検定を行なう。

5.3 解析

まず、上記方法で得られる時刻列より授業開講日の1日分の累積カウント数を図10に示す。事象生起密度関数であるモデル(式(10))のトレンドが指数であるので、図10の累積カウント数の視察からこれに適合するするような観測時間を選ぶ。本稿では1999年10月27日午前8:15から30分間に対し非定常ポアソン性の検定を行なった。また別時刻である1999年11月25日午前8:18から30分間のリクエスト発生時刻列についても同様の解析を行ったがそちらは付録を参照されたい。

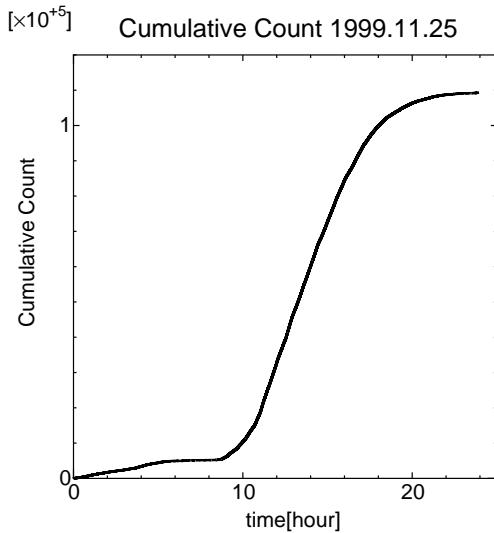


図10. 1999年11月25日の累積カウント数

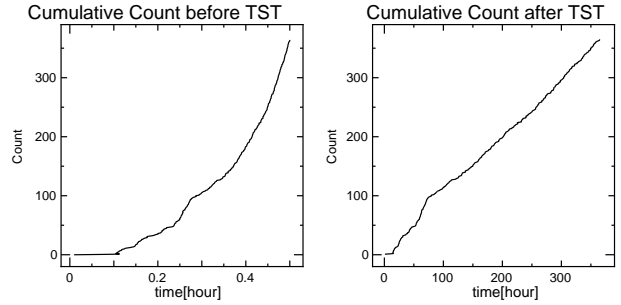


図11. 累積カウント数 (左は時間尺度変換前、右は変換後)

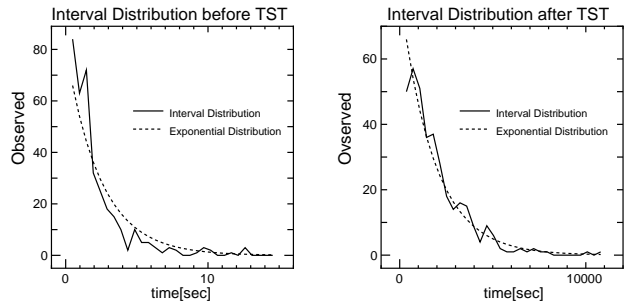


図12. インターバル分布 (左は時間尺度変換前、右は変換後)

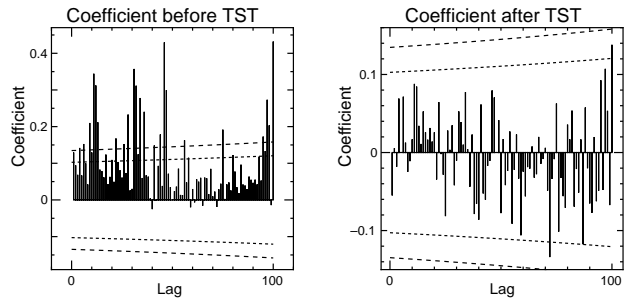


図13. 系列相関係数 (左は時間尺度変換前、右は変換後)

5.4 結果

最尤推定により求められた最尤パラメータは $(\hat{\theta}_1 = 94.607, \hat{\theta}_2 = 6.531)$ となった。累積カウント数を図11に、インターバル分布を図12に、系列相関係数を図13に、および各検定の統計量を表5に示す。

表5. 検定に際する統計量

	時間尺度変換前	時間尺度変換後
事象生起数	365	365
$d (d/9)$	291.301(32.367)	23.466(2.607)
χ^2 適合度	73.47	25.48
5% 有意水準	43.8	43.8
Sherman 統計量 S	0.4679	0.3508
95% 信頼区間	(0.3430, 0.3928)	(0.3430, 0.3928)

5.5 考察

対象とする切り抜いた時刻列の観測原点を0とし累積カウント数を表したものの(図11左)に対し、非定常性を考慮するための時間尺度変換を施した累積カウント数は図11右であり、非定常点過程である実データを定常点過程である1つの実現とみなした。そこで、インターバル分布と χ^2 適合度、生起事象数に関する検定による d の値、系列相関係数、Sherman 統計量を求め時間尺度変換前後での比較を行なう。時間尺度変換前では系列相関係数は長く続く正の相関であり所々で強い相関をもち有意水準内に収まらず、また χ^2 適合度、Sherman 統計量による検定においても有意水準からはずれ帰無仮説は棄却された。それに対し、非定常性を考慮した時間尺度変換後の結果では $d/9$ は時間尺度変換前に比べポアソン過程の理論値1に近い値を取り、系列相関係数、 χ^2 適合度、Sherman 統計量とも有意水準内に収ま

り帰無仮説は棄却されなかった。以上の帰無仮説の棄却検定から、クラスタ化したリクエストの発生時刻列は非定常ポアソン過程である一面が支持できる。

6 おわりに

本報告ではリクエスト発生過程の統計的性質の検討としてリクエストの非定常性を考慮し、非定常過程としての特徴付けの方法とその解析例について述べた。

モデルに関する最尤推定から得られるパラメータをもつ生起事象密度関数の非定常ポアソン過程の時系列は、時間尺度変換により定常ポアソン過程に帰着できることをふまえた上でのモデルの提案を示した。実際のリクエスト発生時刻列を検定する前にシミュレーションを行ない、最尤推定が良く行なわれたことと時間尺度変換により非定常ポアソン過程の見本過程を作成したものに対する定常ポアソン性検定の結果から、上記方法をとることが確かであることが明らかになった。そこで、Proxy サーバ squid のアクセスログから同一ユーザの最初のリクエストから2秒以内を削除しクラスタ化した実際のリクエスト発生時刻列に対し、同様な方法で非定常ポアソン性検定を行なった。非定常性を考慮した時間尺度変換を行なった見本過程では、非定常ポアソン過程であるとの帰無仮説は本稿で挙げた定常ポアソン性検定法では棄却されなかった。これにより、クラスタ化したリクエスト発生時刻列は非定常なポアソン過程の一面を持ち合わせるということがわかった。

しかし、本稿で挙げた定常ポアソン性の検定方法は一部であり、より精密な検定を行なうためにその他の検定方法を用いる場合には文献 [6] を参照されたい。また、授業開講時1日のリクエスト発生密度は図10のようなパターンをもち、本稿で用いた事象生起密度関数モデルでは指数分布に従うもので観測時間が限定されてしまう。一般に事象生起密度関数を完全に表せるようなモデル化は難しいので、様々な形の事象生起密度をもつ非定常ポアソン過程に対しても柔軟に対応できるモデルの検討も有用であると思われる。

参考文献

- [1] 名部雅彦, 馬場健一, 村田正幸, 宮原秀夫, "インターネット・アクセスネットワーク設計のためのWWWトラヒックの分析とモデル化," 信学論 (B-I), Vol. J80-B-I, No. 5, pp. 428-437, Jun 1997.
- [2] 八名和夫, "非定常過程に対するポアソン性検定," 信学論, Vol. J67-A, No. 5, pp. 431-438, 1984.
- [3] D. R. Cox, and Smith, W. L., "The surposition of several atrctly periodic sequences of events," *Biometrika*, Vol. 40, pp. 1-11, 1953.
- [4] 情報理論とその応用学会, "確率過程 - 応用と話題 -," 培風館, 1994.
- [5] 宮沢政清, "確率と確率過程," 近代科学社.

- [6] D. R. Cox, and P. A. W. Lewis, "The statistical analysis of series of events.," Methuen & Co., London, 1966
- [7] 名部雅彦, 村田正幸, 宮原秀夫, "キャッシングを考慮したWWWトラヒック特性の分析とモデル化," 信学論 (B-I), Vol. J81-B-I, No. 5, pp. 325-334, May 1998.
- [8] 小沢利久, 町原文明, 石橋佳介, "マルチメディアトラヒック理論の最新動向," 信学誌, vol. 81, No. 5, pp. 506-515, May 1998.
- [9] W. E. Leland, W. Willinger, M. S. Taqqu, and D. V. Wilson, "On the self-similar nature of Ethernet traffic," SIGCOMM '93, pp. 183-193, 1993.
- [10] S. Pederson, and M. Johnson, "Estimating Model Discrepance," *Technometrics*, no. 3, pp. 305-314, Aug 1990.
- [11] 矢吹道郎監修, 長岡秀行著, "Squid- プロキシサーバの設定と運用 -," テクノプレス, 1998.

付録

1999年11月25日午前8:18から30分間のリクエスト発生時刻列での解析結果を付録として記す。最尤推定により求められた最尤パラメータは ($\hat{\theta}_1 = 34.665520, \hat{\theta}_2 = 8.933512$) となった。このデータにおいて累積カウント数を図14に、インターバル分布を図15に、系列相関係数を図16に、および各検定の統計量を表6に示す。

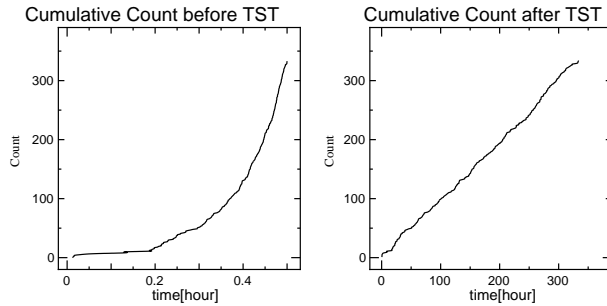


図 14. 累積カウント数 (左は時間尺度変換前、右は変換後)

表 6. 検定に際する統計量

	時間尺度変換前	時間尺度変換後
事象生起数	334	334
$d (d/9)$	425.348(47.260)	6.489(0.721)
χ^2 適合度	102.72	25.27
5% 有意水準	43.8	43.8
Sherman 統計量 S	0.5227	0.3509
95% 信頼区間	(0.3419, 0.3939)	(0.3430, 0.3928)

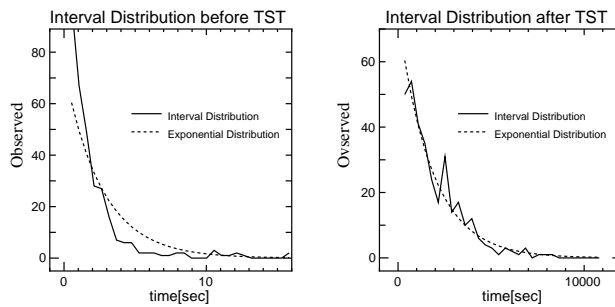


図 15. インターバル分布 (左は時間尺度変換前、右は変換後)

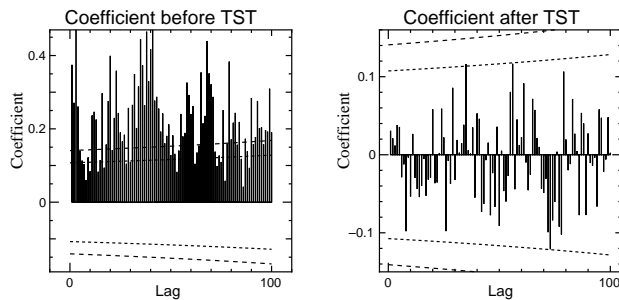


図 16. 系列相関係数 (左は時間尺度変換前、右は変換後)

キーワード.

ネットワークトラフィック、非定常ポアソン性、通信リクエスト

Summary.

**Test for nonstationary Poisson hypothesis of communication request
occurrences in the information network system**

Hideaki Shino Keiichi Kitazawa Masayuki Jibu Kazuo Yana
Department of Electrical Infomatics , Colleague of Engineering , Hosei University

Quantitative grasp of network traffic characteristics is important in designing and operating the information network system. Statistical modeling of the network traffic change enables a large scale computer simulation of a complex network system before actual system development and help efficient design and evaluation of the target system. This paper proposes a method to characterize the statistical properties of the network traffic occurrences. A test for nonstationary Poisson hypothesis using a continuous time scaling is introduced and applied to the real data. The analysis showed that the request occurrences fit well with nonstationary Poisson process.

Keywords.

Network Traffic , Nonstationary Poisson hypothesis, Communication request