確率過程モデルを用いた
ジェット天体の時系列データの
多波長間タイムラグ推定

広島大学理学部物理学科
高エネルギー宇宙・可視赤外天文学研究室
B155406
大間々知輝
主査 植村 誠　副査 岡部 信広

2019 年 2 月

概 要

近年、望遠鏡や衛星の発達により 1 つの天体について多波長の時間変動を観測することが可能になった。ジェット天体の時系列データの中には、2 つの光度曲線にタイムラグを伴って相関しているデータが存在することが分かってきた。また、そのデータの中には畳み込み関数によって広がり（タイムストレッチ）を持っているように見えるデータも存在することが分かっている。これは例えば、中心天体から噴出されるプラズマの到達時間の差が原因で観測されると考えられている。

本研究では確率過程モデルであるオルンシュタイン-ウーレンベック過程を用いて、相関している 2 つの光度曲線のパラメータの推定法を研究した。初めに人工データを用いて、パラメータを変化させた時やデータの誤差が大きい時、データ欠損がある時など様々な条件での推定限界を調べ性能評価を行った。次に、活動銀河核 PKS 1510−089 のガンマ線と電波のデータを推定することで、別の手法を用いて推定された先行研究と同様の 50 日程度のタイムラグと、先行研究では議論されていない 20 日程度のタイムストレッチの推定に成功した。これによりジェットの内部構造について新たな情報を加えて議論することが可能になった。

目 次

第1章　序論

1.1　宇宙ジェット

1.1.1　宇宙ジェットとは

宇宙ジェットとは、活動銀河核、原始星、白色矮星、中性子星、ブラックホールのような中心の天体システムから双方向に吹き出しているプラズマのアウトフローである。宇宙ジェットの特徴をいくつか示す。

中心天体から噴出されたプラズマは、原始星などの低密度の天体においては数十 $\mathrm{km\,s^{-1}}$ から数百 $\mathrm{km\,s^{-1}}$ 程度と低速であるが高密度天体ではかなり高速になり、例えば活動銀河核（AGN）のジェットは光速の 99% にまで加速されることも珍しくない。それほどまでに加速される原因というのは現在のところわかっていないが、放射圧や磁場などが考えられている。次は、収束問題である。観測されている宇宙ジェットは、例えば活動銀河核のジェットなら 1° 以下であることがわかっている。どのようにジェットを収束させているのかというのは理論モデルの要にもなってくる (Shibata et al., 1999)。

宇宙ジェットを観測することによって、天体周辺のガスの情報や磁場の情報を手に入れることができる。降着円盤は中心天体に吸い込まれてしまうのでガスが持っていた情報はほとんどが失われてしまう。しかしジェットは唯一、多くの情報を持って宇宙空間に戻される。これを調べることで天体の構造を議論することができる (Koyama, Minesige, 2007)。

ジェット天体の多くは電波からガンマ線までの広い波長帯で観測できる事がわかっている。ジェットからのシンクロトロン放射は電波から X 線で、降着円盤からの熱的な放射や非熱的な放射が可視光から X 線で観測される。また最近では、フェルミガンマ線宇宙望遠鏡により多くのジェット天体からガンマ線が検出されている。このような理由により、多波長観測は 1 つの波長帯だけでは得られなかった情報が手に入り、特に時間変動の情報からブラックホール周辺の構造、降着円盤やジェットの相互作用、ジェット内部の物理状態など多くの手掛かりが得られる。

1.1.2　マイクロクェーサー

ブラックホールや中性子星などのコンパクト天体を含む連星系の中で、コンパクト天体の重力によって相手の星表面のガスがコンパクト天体に降着し、強い X 線を放射する天体のことを X 線連星という。X 線連星の中で、ジェットが観測される天体は、その構造がクェーサーに似ていることからマイクロクェーサーと呼ばれる事がある。ここでは銀河系内ジェット天体であるマイクロクェーサーの先行研究について 2 つ紹介する。

ブラックホール連星である GRS 1915+105 の X 線光度曲線を図 1.1 に示す。黒点が X 線、赤点が近赤外線、緑点が電波の光度曲線を示している。X 線の 1 分ほどの時間スケールで起きる振幅の変化の後に、強度が大きく落ちていることが確認できる。その後に近赤外線、次に電波の強度が大きくなっていることが確

認できる。これを先行研究では、X 線の強度が弱くなった時にプラズマ放出が起き、シンクロトロン放射を放出するプラズマがジェットの運動とともに広がっているために放射強度が最大となる波長が長くなっていると解釈している (Mirabel, Rodriguez, 1998)。

X 線連星 V404 Cygni の X 線と可視光のタイムラグの研究では、X 線と可視光の同時観測から X 線に比べて可視光が 0.1 秒遅れていることを、後述する相互相関関数を用いて発見した。これを先行研究では、図 1.2 のように、ブラックホール直近で X 線を放射したプラズマが、ジェットの中をブラックホールのシュバルツシルト半径の 1000 倍程度以下だけ移動して可視光を出したと解釈し、この距離がジェットの加速と収束の領域だと結論づけている (Gandhi et al., 2017)。

このように多波長同時観測を行うことでジェットの描像について議論することが可能になる。

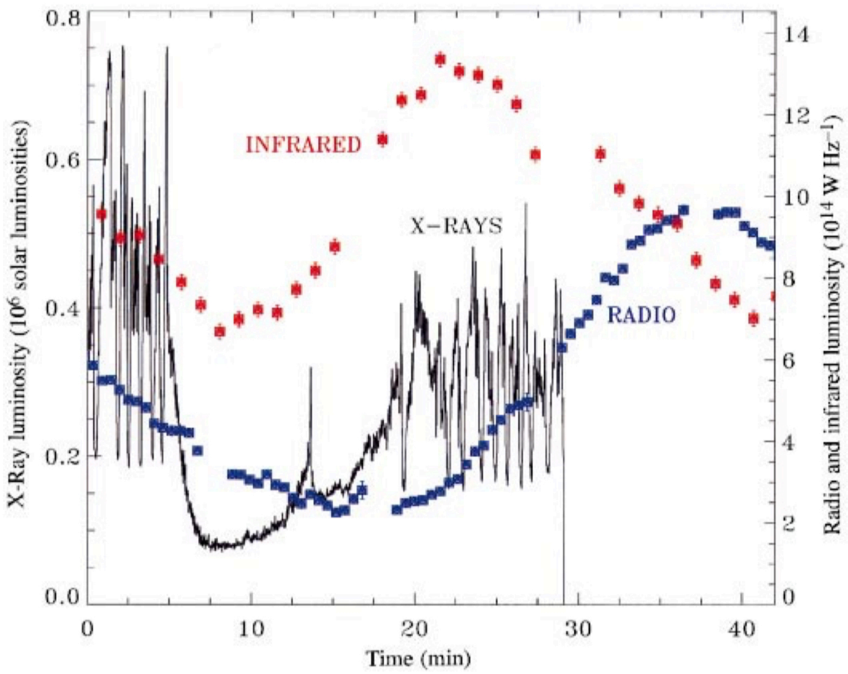


図 1.1: GRS 1915+105NO 電波、近赤外線、X 線の強度の光度曲線 (Mirabel, Rodriguez, 1998)。

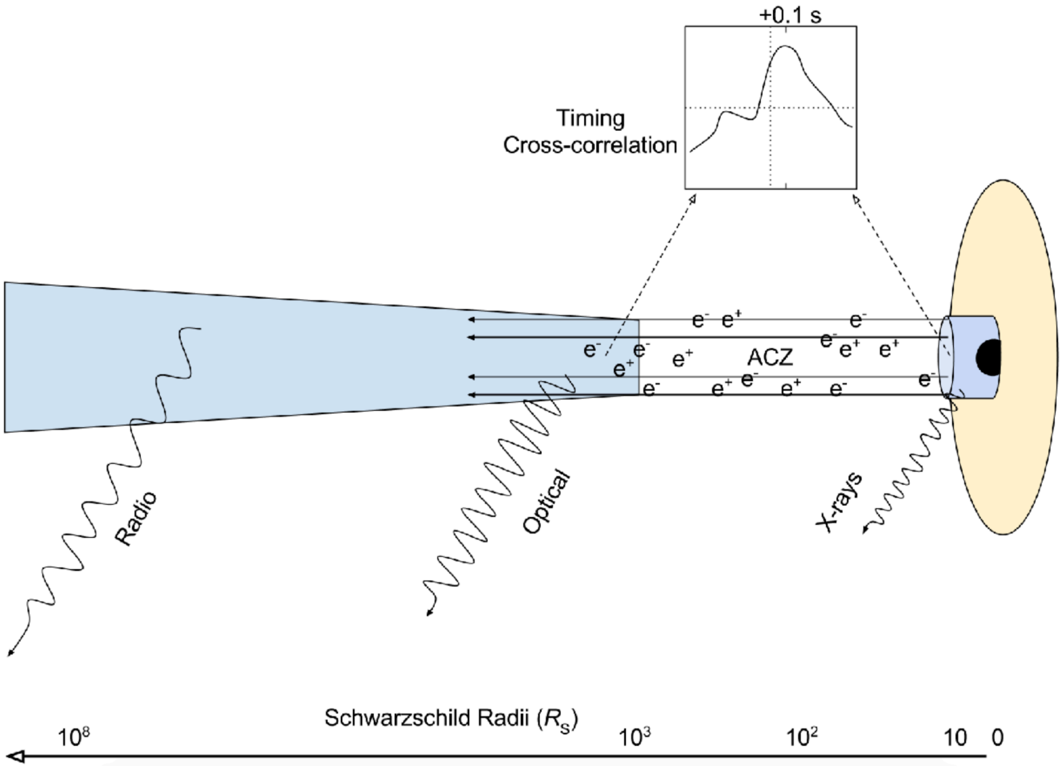


図 1.2: V404 Cygni のイメージ図 (Gandhi et al., 2017)。右上のグラフは相互相関関数である。

1.1.3　活動銀河核

銀河中心の明るさが銀河全体の数倍から数 100 倍明るく、高い活動性を示す天体のことを活動銀河核（AGN）と呼ぶ。AGN の数は銀河全体の 10%ほどを占めていると言われている。その中心には超巨大質量ブラックホールが存在していると言われ、その大きさは $10^{6-8}M_{\odot}$（$M_{\odot}$ は太陽質量を表す）と言われている。AGN を中心天体とするジェットを正面から観測した天体を「ブレーザー」と言う。ブレーザーの特徴として激しい時間変動とガンマ線放射がある。他にも、超光速現象と呼ばれるジェットが光速を超えて移動して見える現象がある。ブレーザーは、他の AGN よりも降着円盤、ダストトーラス、銀河からの放射が小さく、ジェットの内部構造を探るのに適した天体である。図 1.3 に AGN の想像図を示す。降着円盤やトーラスが中心天体の周りに位置し、ジェットが噴き出している様子が描かれている。

ブレーザーである PKS 1510−089 の先行研究を紹介する。図 1.4 にガンマ線と電波の光度曲線の図を示す。通常タイムラグの推定には後述する Discrete Correlation Function(DCF) を用いるが、この天体では電波フレアの形状がガンマ線に比べて鈍く DCF では値を推定できなかったため、フレアの極大時間を比較することでタイムラグの推定を行ない、54 日のタイムラグがあると結論づけている (Beaklini et al., 2017)。本論文ではこの鈍りをタイムストレッチと呼び、畳み込み関数の幅として定義している。詳しくは 2.2 章に書いている。

AGN の内部構造の情報を得る手段として reverberation mapping がある。これは可視連続光と輝線光のタイムラグから輝線光の散乱場所を推定する手法である。図 1.5 のように、AGN の中心付近から放射され

た可視光の連続光は直接我々に届くが、周辺の物質で散乱された輝線光は遅れて届く。例として連続光と輝線光にタイムラグが生じている光度曲線を図 1.6 に示す。これは上から紫外線の連続光、可視光の連続光、ライマン α 輝線、バルマー β 輝線の光度曲線を示している。この光度曲線の情報からタイムラグを推定する事で散乱物質の場所が特定できる。

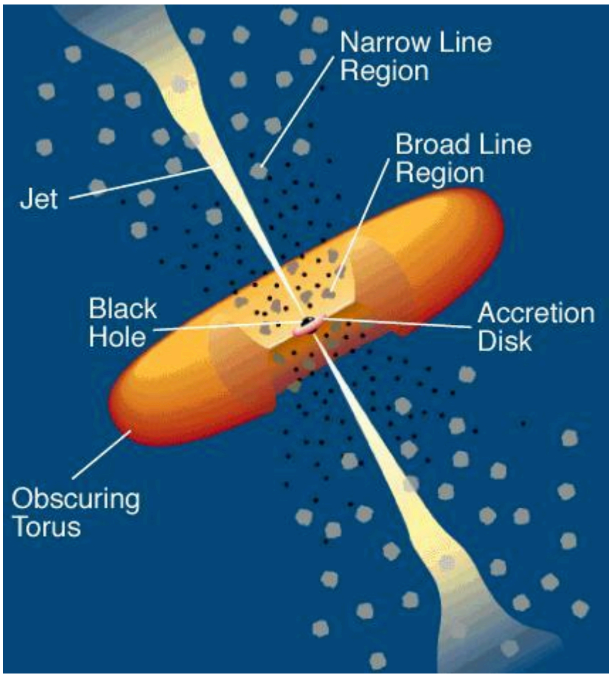


図 1.3: AGN の想像図 (Urry, Padovani, 1995)

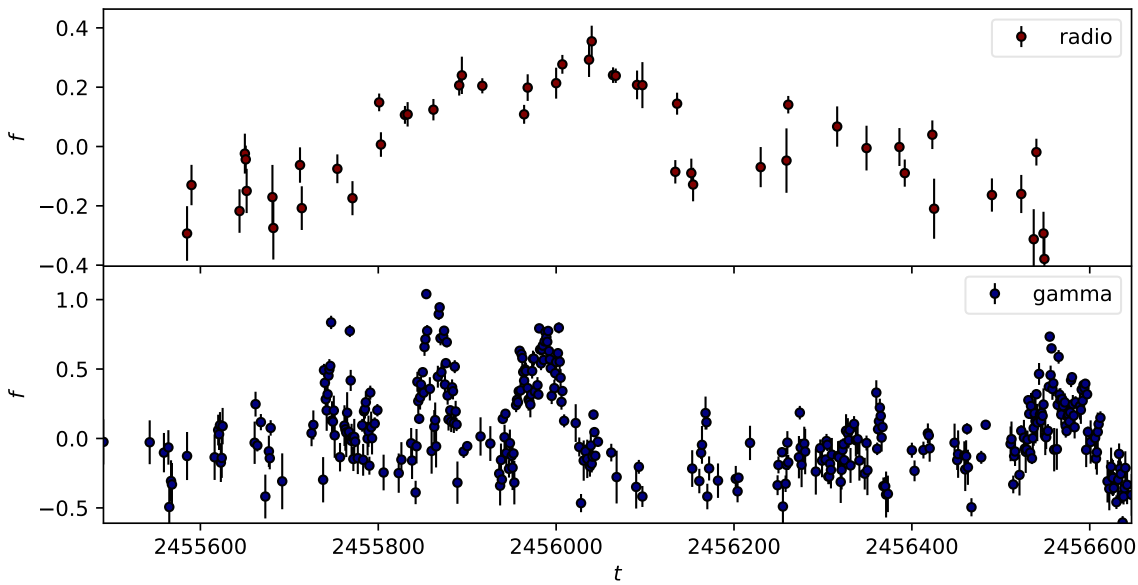


図 1.4: PKS 1510−089 の光度曲線

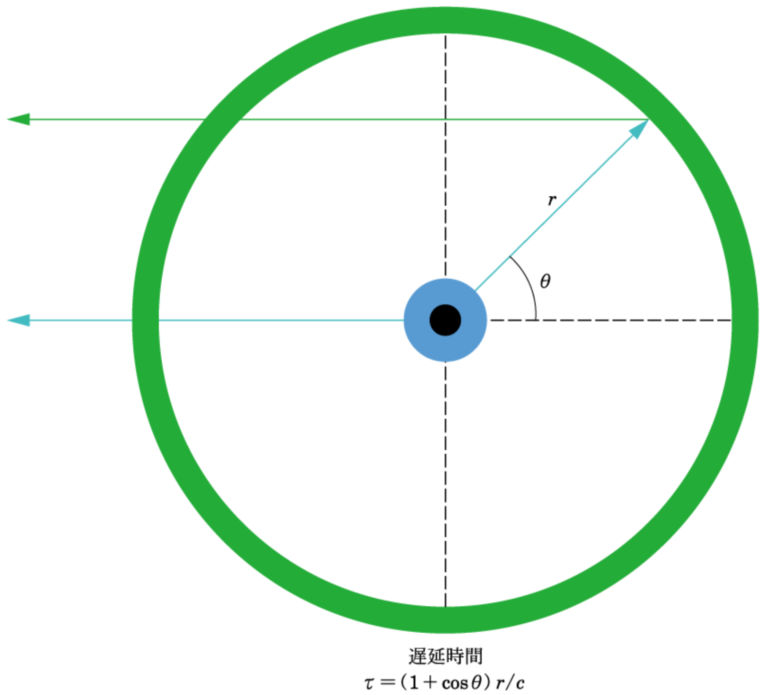


図 1.5: reverberation mapping の模式図（天文学辞典より http://astro-dic.jp/reverberation-mapping/）

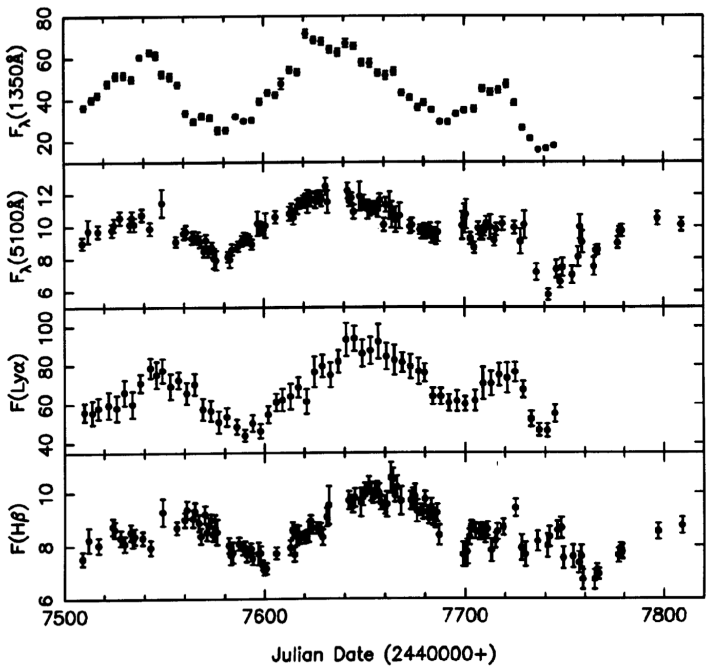


図 1.6: NGC 5548 の多波長同時観測した光度曲線 (Peterson, 1993)。

1.2 相関関数

タイムラグがある 2 つの時系列データにおいて、従来、タイムラグの推定には相関係数を用いることが多かった。ここでは、相関係数について説明し、その問題点について述べる。

相関係数とは、2 つの変数の相関関係の強さを測る尺度である。確率変数 x、y の相関係数 r は式 (1.1) で表される。

$$r = \frac{s_{xy}}{s_x s_y} \tag{1.1}$$

ここで s_x、s_y は x と y の標準偏差であり、s_{xy} は共分散であり

$$s_x = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n-1}} \tag{1.2}$$

$$s_y = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \overline{y})^2}{n-1}} \tag{1.3}$$

$$s_{xy} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y}))}{n-1} \tag{1.4}$$

と表される。

相互相関関数

相互相関関数とは 2 つの関数の類似性を測る関数である。分散が 1、平均が 0 で標準化した 2 つの関数を $f(t)$、$g(t)$ とすると、2 つ関数の畳み込みで表される相互相関関数 $z(t_{\mathrm{lag}})$ は式 (1.5) で表される。

$$z(t_{\mathrm{lag}}) = \sum_t f(t - t_{\mathrm{lag}})g(t) \tag{1.5}$$

$f(t)$ と $g(t)$ が同じ関数であれば、自己相関関数と呼ばれる。相関係数の誤差は、統計的検定で表現される。例として $t_{\mathrm{lag}} = -10$ の相関が最も強い光度曲線を図 1.7 に、この光度曲線を相関係数を用いて解析した時の結果を図 1.8 に示す。図 1.8 では $t_{\mathrm{lag}} = -10$ で相関係数が最も大きくなっており、この結果から、タイムラグが -10 である事が分かる。このように 2 つの時系列データが同時刻で取得されている場合は相互相関関数によってタイムラグの推定が可能になる。

Discrete Correlation Function

相互相関関数では、定義から等間隔のデータしか扱うことかできない。Discrete Correlation Function（DCF）は、相互相関関数を拡張し、等間隔ではないデータを扱うための手法である。

DCF の計算は以下の通りである。全ての 2 つの配列 a_i と b_i の組に対し、unbinned discrete correlation（UDCF_{ij}）を式 (1.6) のように計算する。

$$\mathrm{UDCF}_{ij} = \frac{(a_i - \overline{a})(b_j - \overline{b})}{\sqrt{(\sigma_a^2 - e_a^2)(\sigma_b^2 - e_b^2)}} \tag{1.6}$$

ここで σ_a と σ_b はそれぞれ a_i と b_i の標準偏差であり、e_a と e_b はそれぞれ測定誤差である。得たい DCF の時間ビンの幅を Δt_{lag} とすれば、式 (1.6) で計算した UDCF を時間ビンでまとめて、DCF(t_{lag}) は式 (1.7) のように表される。

$$\mathrm{DCF}(t_{\mathrm{lag}}) = \frac{1}{M}\mathrm{UDCF}_{ij} \text{ for } t_{\mathrm{lag}} - \frac{\Delta t_{\mathrm{lag}}}{2} \leq t_i - t_j < t_{\mathrm{lag}} + \frac{\Delta t_{\mathrm{lag}}}{2} \tag{1.7}$$

ここで M は時間ビンにある UDCF の組の数である。誤差は同じ時間ビンに入る $\mathrm{UDCF_{ij}}$ の標準偏差として、式 (1.8) で表される (Edelson, Krolik, 1988)。

$$\sigma_{\mathrm{DCF}}(t_{\mathrm{lag}}) = \frac{1}{M-1}\left\{\sum\left[\mathrm{UDCF}_{ij} - \mathrm{DCF}(t_{\mathrm{lag}})\right]^2\right\}^{1/2} \tag{1.8}$$

図 1.9 に DCF の例を示す。この時解析した光度曲線は図 1.7 を用いた。時間ビンの幅は 2 である。相関係数は定義上 -1 から $+1$ の間の値を持つ。しかし $t_{\mathrm{lag}} = -80$ 付近では誤差も含めて -1 から 1 の間に入っていないことが確認できる。このように、DCF はデータの測定誤差次第では相関係数の定義に当てはまらない挙動をすることがある。

図 1.8 や図 1.9 から分かるように、相関係数や DCF を用いた推定では、相関係数の誤差は統計的検定から与えられるが、タイムラグの不定性は評価が難しい。また 1.1.3 章で述べた PKS 1510−089 のデータのように、異なる波長間でタイムスケールが異なる場合は、相関を低く評価してしまう可能性がある。また、その定義式から観測間隔よりも小さいタイムラグの推定はできない。

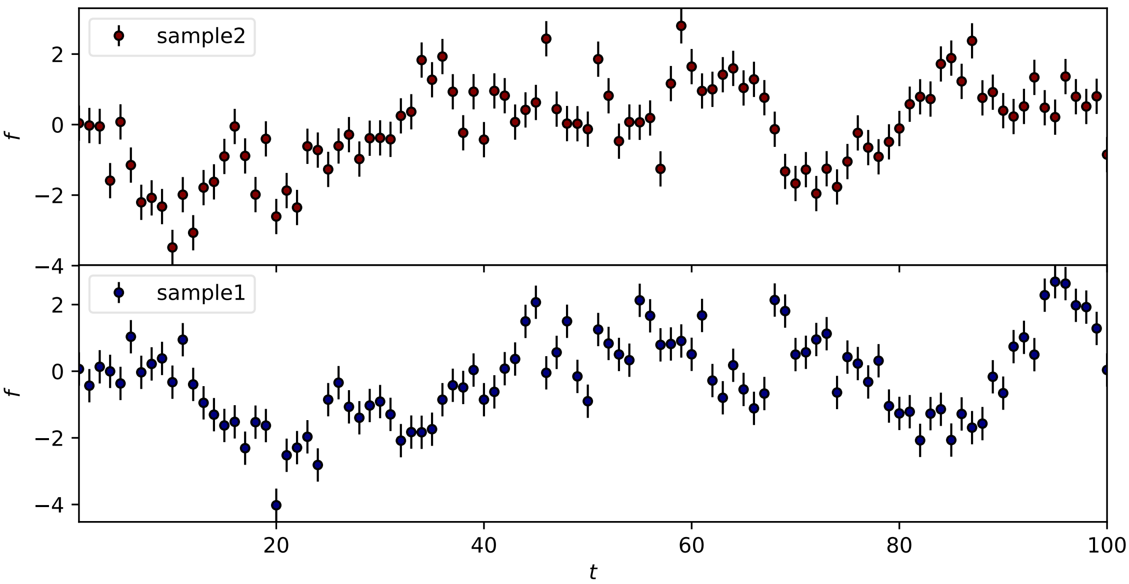


図 1.7: -10 のタイムラグがある光度曲線。

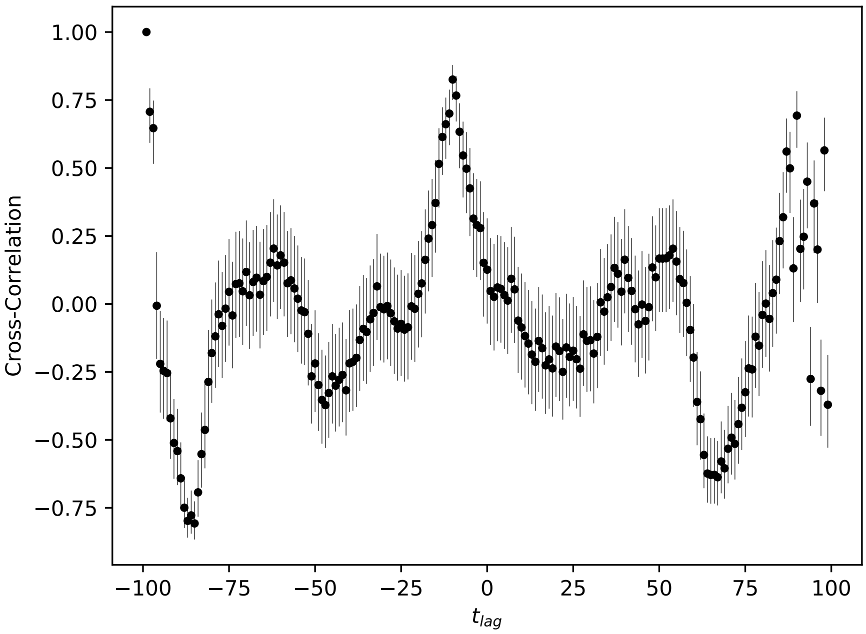


図 1.8: 相互相関関数の例。$t_{\mathrm{lag}} = -10$ に最も強い相関がある。

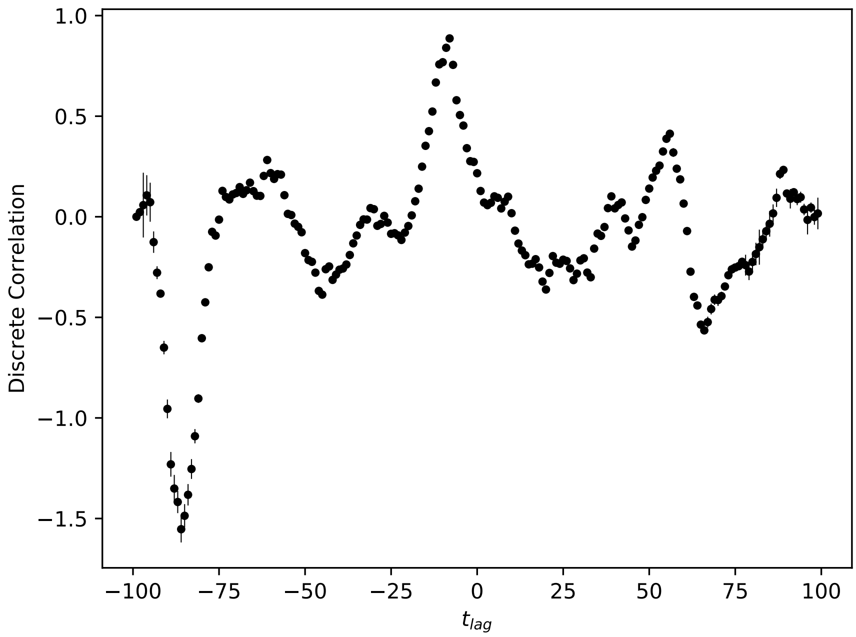


図 1.9: DCF の例。$t_{\mathrm{lag}} = -10$ に最も強い相関がある。

1.3 本研究の目的

多波長同時観測されたジェット天体の中には、タイムラグを伴って相関しているように見えるデータが多く存在し、さらに、PKS 1510−089 のようにタイムラグだけでなくタイムストレッチもあるように見えるデータも存在することが分かっている。タイムラグやタイムストレッチからは放射源の場所や大きさ、天体の構造などの手がかりを得る事ができるが、タイムラグの推定は従来の相関係数を用いる方法が依然として主流であり、タイムストレッチについては推定法やそれを用いた議論が確立していない。そこで本研究では、確率過程モデルであるオルンシュタイン-ウーレンベック過程（OU 過程）を用いて光度曲線をモデル化し、その光度曲線を用いてタイムラグとタイムストレッチを推定する。OU 過程や応答関数のパラメータの最適解を統計手法であるマルコフ連鎖モンテカルロ法（MCMC）を用いて推定する。OU 過程と応答関数のパラメータを MCMC を用いて推定する方法は、従来の方法では曖昧になっていた誤差の問題や、季節や天気などの影響によるデータ欠損についての問題も解決することが期待される。本研究は、タイムラグとタイムストレッチの推定をすることでジェットの内部構造について新たな情報を加えて議論することを目的とする。

本論文は以下のように構成されている。2 章では確率過程モデルであるガウス過程と、その特殊な場合である OU 過程について説明する。3 章では OU 過程を用いたタイムラグとタイムストレッチの推定について、最初に人工データを用いた実験を行い、次に実際のデータへの応用例として PKS 1510−089 のデータ解析を行いその結果を報告する。4 章では前章の実験結果に関する考察と、PKS 1510−089 のデータの解析結果から得られる宇宙物理学的な示唆について議論する。

第 2 章　解析手法

2.1　ガウス過程

得られているデータをベクトル $\boldsymbol{t}$ で表す。ガウス過程は一般的には時系列データの解析に限らず、他の種類のデータにも応用することができるが、ここでは時系列データである光度曲線を想定してガウス過程について説明する。

N 個の光度 $\boldsymbol{s}$ の同時分布が多変量ガウス分布に従う時、これをガウス過程に従うという。この時、$\boldsymbol{s}$ は以下の式で表される

$$\boldsymbol{s} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{K}) \tag{2.1}$$

ここで $\boldsymbol{\mu}$ は平均ベクトル、$\boldsymbol{K}$ は後に示すカーネル関数で表される共分散行列である。なお、光度の平均をあらかじめ引いておくことで $\boldsymbol{\mu} = \boldsymbol{0}$ にすることができるので、煩雑さを避けるため、多くの場合で

$$\boldsymbol{s} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{K}) \tag{2.2}$$

として扱われる。ガウス過程において重要なのは N 個の変数 s_1、...、s_n の同時分布が平均と共分散の 2 次の統計量で完全に記述される点である。

データにガウス過程を当てはめることをガウス過程回帰すると言う。以下、ガウス過程を回帰問題に適用していく。式 (2.3) のように観測されるデータ $\boldsymbol{y}$ に誤差が含まれているとする。

$$y_i = s_i + \varepsilon_i \ (i = 1, 2, \cdots, \mathrm{N}) \tag{2.3}$$

ここで $s_i = s(t_i)$ であり、ε_i は i 番目の観測値に加えられる誤差であり、これは平均 0、分散 σ_s^2 の正規分布に従う。誤差は各データに対して独立に決まるので、$\boldsymbol{s}$ が与えられた時の $\boldsymbol{y}$ の同時分布は式 (2.4) のように表される。

$$p(\boldsymbol{y}|\boldsymbol{s}) = \mathcal{N}(\boldsymbol{y}|\boldsymbol{s}, \sigma^2 \boldsymbol{I}_N) \propto |\sigma_s^2 \boldsymbol{I}_N|^{-1/2} \exp\left(-\frac{(\boldsymbol{y}-\boldsymbol{s})^T \sigma_s^2 \boldsymbol{I}_N^{-1} (\boldsymbol{y}-\boldsymbol{s})}{2}\right) \tag{2.4}$$

$\boldsymbol{I}_\mathrm{N}$ は $\mathrm{N} \times \mathrm{N}$ の単位行列である。また、$\boldsymbol{s}$ の事前分布 $p(\boldsymbol{s})$ は、平均が $\boldsymbol{0}$、共分散行列が K の多変量正規分布なので

$$p(\boldsymbol{s}) = \mathcal{N}(\boldsymbol{s}|\boldsymbol{0}, \boldsymbol{K}) \propto |\boldsymbol{K}|^{-1/2} \exp\left(-\frac{\boldsymbol{s}^T \boldsymbol{K}^{-1} \boldsymbol{s}}{2}\right) \tag{2.5}$$

で与えられる。時刻 $\boldsymbol{t}$ で光度曲線データが得られた時、$\boldsymbol{y}$ が得られる確率は、事後確率 $p(\boldsymbol{y}|\boldsymbol{s})p(\boldsymbol{s})$ に対して $\boldsymbol{s}$ についての周辺積分を行えばよく、

$$p(\boldsymbol{y}) = \int p(\boldsymbol{y}|\boldsymbol{s}) p(\boldsymbol{s}) d\boldsymbol{s} = \mathcal{N}(\boldsymbol{y}|\boldsymbol{0}, \mathbf{C}) \tag{2.6}$$

のように求めることができる。ここで共分散行列 $\boldsymbol{C}$ は要素

$$\boldsymbol{C}(t_i, t_j) = k(t_i, t_j) + \sigma_s^2 \delta_{ij} \tag{2.7}$$

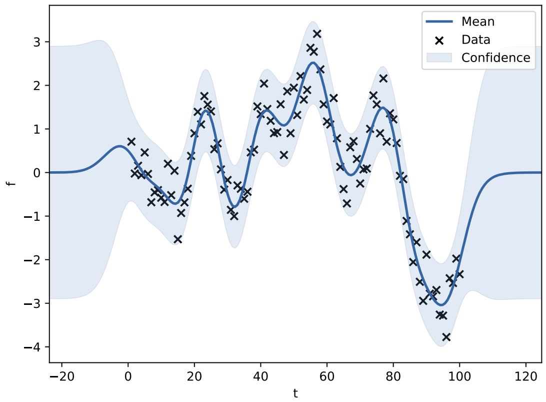


図 2.1: ガウス過程から生成したサンプルとガウス過程で回帰した光度曲線。$\sigma = 1$、$\tau = 5$ である。

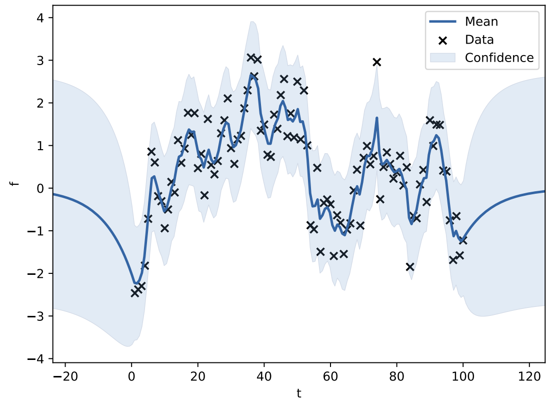


図 2.2: OU 過程から生成したサンプルと OU 過程で回帰した光度曲線。$\sigma = 1$、$\tau = 5$ である。

である。式 (2.6) から尤度関数を定義する事ができ

$$p(\boldsymbol{y}) \propto \mathcal{L}(\boldsymbol{y}) \equiv |\boldsymbol{C}|^{\frac{1}{2}} \exp\left(-\frac{\boldsymbol{y}^T C^{-1} \boldsymbol{y}}{2}\right) \tag{2.8}$$

と表される。

ここで、$k(t_i, t_j)$ はカーネル関数であり、例えば以下の関数系のものがよく使われてる。

$$k(t_i, t_j) = \sigma \exp(-\frac{|t_i - t_j|^2}{2\tau^2}) \tag{2.9}$$

$$k(t_i, t_j) = \sigma \exp(-\frac{|t_i - t_j|}{\tau}) \tag{2.10}$$

式 (2.9) は Radial Basis Function(RBF) カーネルと呼ばれ、一般的なガウス過程でよく使われるカーネルである。式 (2.10) を用いればオルンシュタイン-ウーレンベック過程（OU 過程）と呼ばれる確率過程になる。図 2.1 には RBF カーネルを使ったガウス過程で生成したサンプルとガウス過程で回帰した光度曲線の例、図 2.2 には OU 過程から生成したサンプルと OU 過程で回帰した光度曲線の例を示している。これらのパラメータどちらも $\sigma = 1$、$\tau = 5$ である。青い曲線は確率分布の中央値、幅を持った区間は 95%信頼区間を示している。この信頼区間とは、ガウス過程に従うのであれば 95%の確率で入力に対してこの区間に入ることを示しており、最も確率が高いのが中央値である。RBF カーネルのガウス過程を用いた回帰では、データに対し比較的緩やかに変動しているのに対し、OU 過程を用いた回帰ではデータに対して敏感に変動していることが確認できる。

以上のように、ガウス過程回帰ではカーネル関数のパラメータ（σ、τ）、及び、測定ノイズである正規分布の分散（σ_s^2）がモデルのパラメータとなり、これらの変数が与えられれば共分散行列 $\boldsymbol{C}$ が定義され、データの尤度の式 2.8 が計算される。尤度を最大にするモデルパラメータを推定するのがガウス過程回帰である。

次に予測分布を求める。そのためには $p(t_{\mathrm{N}+1}|\boldsymbol{t})$ を求める必要がある。式 (2.6) より同時分布は以下のように表される。

$$p(\boldsymbol{y}_{\mathrm{N}+1}) = \mathcal{N}(\boldsymbol{y}_{\mathrm{N}+1}|\boldsymbol{0}, \boldsymbol{C}_{\mathrm{N}+1}) \tag{2.11}$$

ここで、$C_{\mathrm{N}+1}$ は、$(\mathrm{N}+1)\times(\mathrm{N}+1)$ の共分散行列であり、

$$C_{\mathrm{N}+1} = \begin{pmatrix} \boldsymbol{C}_{\mathrm{N}} & \boldsymbol{k} \\ \boldsymbol{k}^T & c \end{pmatrix} \tag{2.12}$$

である。C_N は要素が式 (2.7) であるような $\mathrm{N}\times\mathrm{N}$ の共分散行列、$\boldsymbol{k}$ は要素 $k(t_i, t_{\mathrm{N}+1})$ $(i = 1、\cdots、\mathrm{N})$ をもつベクトル、$c = k(t_{\mathrm{N}+1}, t_{\mathrm{N}+1}) + \sigma_s^2$ である。この時、条件付き確率は、次に示す平均と分散をもつガウス分布になることがわかる。

$$m(t_{\mathrm{N}+1}) = \boldsymbol{k}^T \boldsymbol{C}_{\mathrm{N}}^{-1} \boldsymbol{y} \tag{2.13}$$

$$\sigma^2(t_{\mathrm{N}+1}) = c - \boldsymbol{k}^T \boldsymbol{C}_{\mathrm{N}}^{-1} \boldsymbol{k} \tag{2.14}$$

つまり、新たな入力 $t_{\mathrm{N}+1}$ に対する新しい出力は以下のように表される。

$$p(\boldsymbol{y}_{\mathrm{N}+1}) = \mathcal{N}(m(t_{\mathrm{N}+1}), \sigma^2(t_{\mathrm{N}+1})) \tag{2.15}$$

このように、ガウス過程を用いることで得られたデータに対する確率分布が得られ、データが得られていない区間でも確率分布として予測することができるのである (Bishop, 2012)。

2.2 解析ソフト: JAVELIN

JAVELIN はガウス過程を用いてタイムラグを推定する Python のモジュールである。1.1.3 章で紹介した AGN の連続光と輝線光の光度曲線間にあるタイムラグから AGN の周辺構造を探る「reverberation mapping」と呼ばれる解析手法のために開発された。本研究ではこの JAVELIN を一般的なタイムラグ推定のツールとして用いる。

JAVELIN は次のように 2 つの光度曲線をモデル化する。以下、OU 過程から生成された光度曲線をサンプル 1 と呼び、サンプル 1 と式 (2.18) の積分で生成される、タイムラグをもつ光度曲線をサンプル 2 と呼ぶ。ある天体の光度曲線が OU 過程に従うサンプルだとするとサンプル 1 の光度曲線は次のように表される。

$$\boldsymbol{s}_1(\boldsymbol{t}) = \mathcal{N}(\overline{\boldsymbol{c}}, \boldsymbol{K}) \tag{2.16}$$

ここで $\overline{c}$ はデータの平均、$\boldsymbol{K}$ は共分散行列で、その要素 $k(t_i, t_j)$ は式 (2.10) で表されるカーネル関数、すなわち OU 過程である。サンプル 2 は式 (2.17) に示す畳み込み関数によって式 (2.18) のように表現される。JAVELIN では畳み込み関数はトップハット関数を用いている。

$$\Psi(t) = \frac{A}{w} \text{ for } t_{\text{lag}} - \frac{w}{2} \leq t < t_{\text{lag}} + \frac{w}{2} \tag{2.17}$$

$$s_2(t) = \int \Psi(t - t') s_1(t') dt' \tag{2.18}$$

ここで w、A、はそれぞれトップハット関数の幅と振幅である。t_{lag} は任意のタイムラグである。

これらの光度曲線で構成されたベクトルを $\boldsymbol{s}$ で表し、共分散行列を $\boldsymbol{S}$ で表すと、ガウス過程の定義からこれらのサンプルが得られる確率は次のように表される。

$$p(\boldsymbol{s}) \propto |\boldsymbol{S}|^{-\frac{1}{2}} \exp\left(-\frac{\boldsymbol{s}^T \boldsymbol{S}^{-1} \boldsymbol{s}}{2}\right) \tag{2.19}$$

$\boldsymbol{S}$ の成分はサンプル 1 同士の共分散 $k_{1,1}(t_i, t_j)$ については指数カーネルで与えられるがサンプル 1 とサンプル 2 の共分散 $k_{2,1}(t_i, t_j)$、サンプル 2 同士の共分散 $k_{2,2}(t_i, t_j)$ は、それぞれ以下のように与えられる。

$$k_{2,1}(t_i, t_j) = \int dt' \Psi(t_i - t') k_{1,1}(t', t_j) \tag{2.20}$$

$$k_{2,2}(t_i, t_j) = \int dt' dt'' \Psi(t_i - t') \Psi(t_j - t'') k_{1,1}(t', t'') \tag{2.21}$$

Ψ がトップハット関数の場合、これらの積分は解析的に計算する事ができる。これらを用いると観測データ $\boldsymbol{y}$ は次の式で表される。

$$\boldsymbol{y} = \boldsymbol{s} + \boldsymbol{\varepsilon} + \boldsymbol{L}\boldsymbol{q} \tag{2.22}$$

$\boldsymbol{L}$ は線形成分のための応答関数、$\boldsymbol{q}$ は線形定数であり、これらは光度曲線の線形成分を除くためにある。また、測定誤差である ε はデータの測定誤差を標準偏差とする正規分布を考える。以上より、共分散行列 $\boldsymbol{S}$、線形定数 $\boldsymbol{q}$ が与えられた時の観測データ $\boldsymbol{y}$ の確率は（尤度：$\mathcal{L}$）は 2.1 章と同様、事後確率を s と q で周辺積分することで以下のように得られる。

$$p(\boldsymbol{y}) \propto \mathcal{L}(\boldsymbol{y}) \equiv |\boldsymbol{C}|^{-\frac{1}{2}} |\boldsymbol{L}^T \boldsymbol{C}^{-1} \boldsymbol{L}|^{-\frac{1}{2}} \exp\left(-\frac{\boldsymbol{y}^T \boldsymbol{C}_\perp^{-1} \boldsymbol{y}}{2}\right) \tag{2.23}$$

ここで $\boldsymbol{C}_\perp^{-1} = \boldsymbol{C}^{-1} - \boldsymbol{C}^{-1}\boldsymbol{L}\boldsymbol{C}_q\boldsymbol{L}^T\boldsymbol{C}^{-1}$ であり、$\boldsymbol{C}_q = (\boldsymbol{L}^T\boldsymbol{C}^{-1}\boldsymbol{L})^{-1}$ である。この尤度関数が最大の時のパラメータが最適なパラメータということになる。JAVELIN のパラメータは、OU 過程のカーネルパラメータである σ と τ、畳み込み関数であるトップハット関数のタイムラグ $t_{\rm lag}$、幅 w 及び高さ A である。この時、マルコフ連鎖モンテカルロ法（MCMC）と呼ばれる、乱数を用いてパラメータの事後確率分布を求める統計的手法を用いて尤度関数を最大にするパラメータを求めていく (Zu et al., 2011)。

第3章　OU過程を用いたタイムラグ推定

3.1　人工データを用いた性能評価

実際の天体を解析する前に OU 過程を用いたタイムラグ推定の性能や限界を調べる。モデルのパラメータは、OU 過程の分散パラメータ σ、時間スケール τ とタイムラグ t_{lag}、畳み込み関数のタイムストレッチ w と振幅 A の 5 つである。また、人工データのパラメータはサンプリング間隔、観測基線長と誤差の 3 つである。人工データはまず時刻を決め、次にモデルのパラメータを与えて OU 過程に従うサンプルを生成する。そのサンプルに応答関数を畳み込むことでタイムラグとタイムストレッチのある別の光度曲線を生成する。最後に、それぞれ別の正規乱数を 2 つのサンプルに足すことで測定誤差を含んだ人工データを生成した。パラメータの既定値として、σ は 1、τ は 5、t_{lag} は 20、w は 2、畳み込み関数の振幅は 1 とする。データ数は 100 点、3.1.5 節で誤差の影響を調べる実験以外は無視できるほど小さい場合を考えるので、誤差の標準偏差は 0.000005 を既定値とする。この時の光度曲線の例を図 3.1 に示す。

各実験でこれらの値を変えて推定の限界を調べる。σ、A は光度や等級など天体の明るさを示す量が想定されるが、モデルは他の量についても扱うことができるので本実験では無次元量とする。本実験ではデータ間隔を 1 として、時間に関する物理量の単位は定義しない。

3.1.1 節では推定結果について OU 過程のパラメータである σ と τ の実験全体での傾向について述べる。3.1.2 節では、2 つの時系列データ間の応答関数のパラメータである t_{lag} と w を変化させ関係性を調べる。3.1.3 節では相関関数や DCF を用いた推定では不可能な観測間隔よりも小さい t_{lag} の推定を行う。3.1.4 節ではデータ欠損があるときの人工データで、欠損数の限界について調べる。3.1.5 節ではデータの誤差がデータに対してどのくらいの大きさなら推定に成功するのか調べる。

以下、1σ（68%信頼区間）内に真値があるとき推定に成功しているとし、明確な大局解がない時、局所解が複数ある時、1σ 内に真値がない時は推定に失敗しているとする。

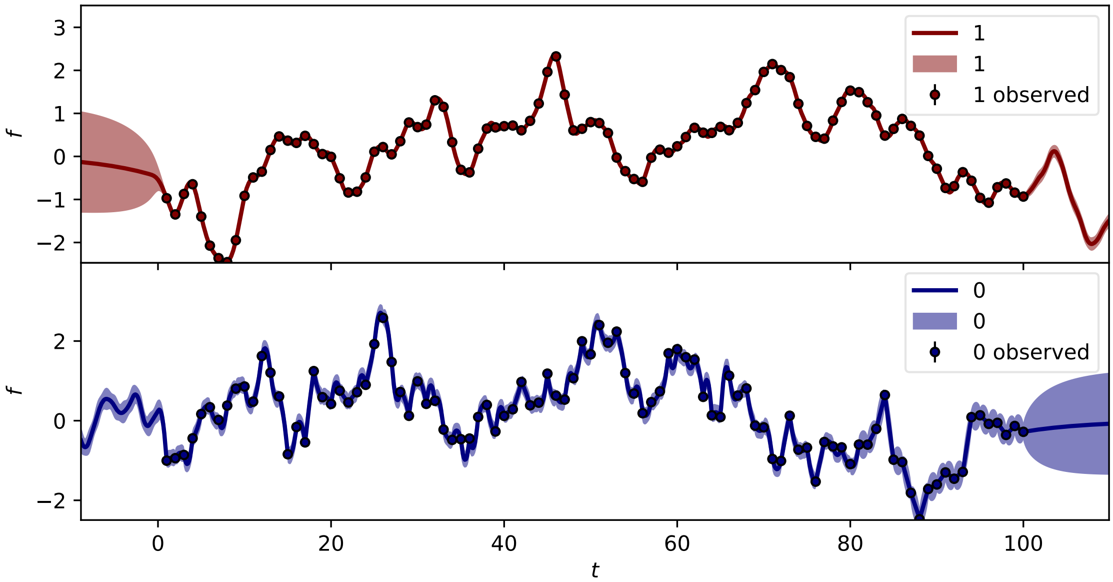


図 3.1: t_{lag} が 20、w が 2 の時の光度曲線。横軸が時間、縦軸が光度を示している。

3.1.1 σ と τ の推定について

実験の結果に入る前に、σ と τ の推定結果について、実験全体での傾向を述べておく。これらのパラメータは実験全体を通して、ほとんどの場合で事後確率の 95%信頼区間内に真値が含まれるものの、分布全体の幅が広く、真値が分布の中心から大きく外れる場合も多く見られた。これらの例として真値が分布の中心付近にある時の事後確率を図 3.2 に、真値が分布の中心から大きく外れている時の事後確率を図 3.3 に示す。σ と τ の真値はそれぞれ $\ln 1 = 0$ と $\ln 5 = 1.6$ である。図 3.2 では、σ と τ の分布がそれぞれ真値である $\ln 1 = 0$ と $\ln 5 = 1.6$ 付近に極大をもち、推定に成功していることが確認できる。図 3.3 では σ と τ の分布は値が大きい側に裾を持っており、このような場合に真値が分布の中心から外れることが多い。

OU 過程のパラメータである σ と τ の推定値は強く相関することが知られており、特に τ が観測期間に対して短い時、モデルが縮退してしまい τ が強く制約されないことがある (Kozłowski, 2017)。従って、モデルの特性上、σ と τ は推定が特に難しいパラメータであることが言える。しかし、図 3.3 からも分かるように、σ と τ の推定に失敗する場合でも、t_{lag} や w の推定は成功することが分かった。これは σ や τ の分布の中心が真値付近にない時でも両方のサンプルの光度曲線のモデルは同一として推定をするので、t_{lag} や w の推定には影響がないことが考えられる。また A は、t_{lag} の真値付近に事前分布を置く事でほとんどの場合、分布の中心が真値付近に存在するので、このパラメータも注目しないことにする。このような事実を踏まえて、以降の実験では σ、τ と A には注目せず t_{lag} と w のみに注目することにする。

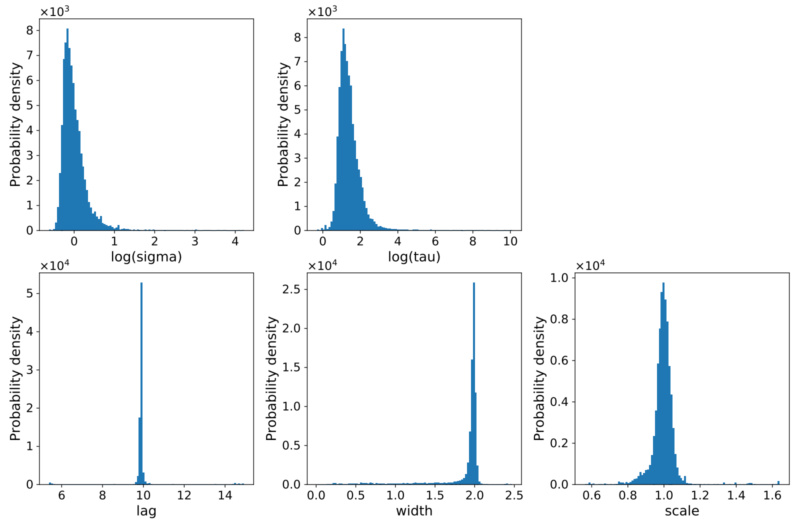


図 3.2: t_{lag} が 10、w が 2 の時の事後確率。t_{lag} には 15 から 25 の事前分布を置いている。横軸は上段左から σ、τ、下段左から t_{lag}、w、A である。縦軸は確率密度である。σ、τ、t_{lag}、w、A の分布の中心が真値 $\ln\sigma = 0.0$、$\ln\tau = 1.6$、$t_{\mathrm{lag}} = 10$、$w = 2.0$、$A = 1.0$ 付近に存在していることが確認できる。

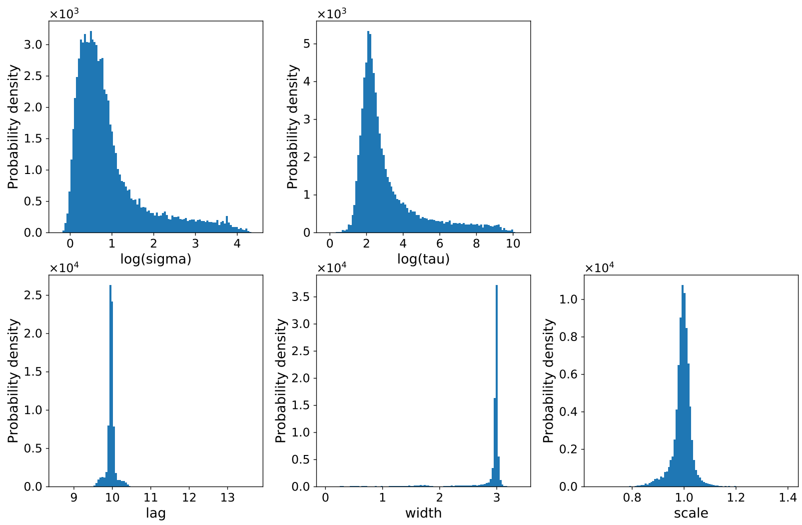


図 3.3: t_{lag} が 10、w が 3 の時の事後確率。t_{lag} には 15 から 25 の事前分布を置いている。横軸は上段左から σ、τ、下段左から t_{lag}、w、A である。縦軸は確率密度である。σ、τ は値が大きい側に裾を持っており真値が分布の中心にないことが確認できる。しかし t_{lag}、w、A は分布の中心が真値 $t_{\mathrm{lag}} = 10$、$w = 2.0$、$A = 1.0$ 付近に存在していることが確認できる。

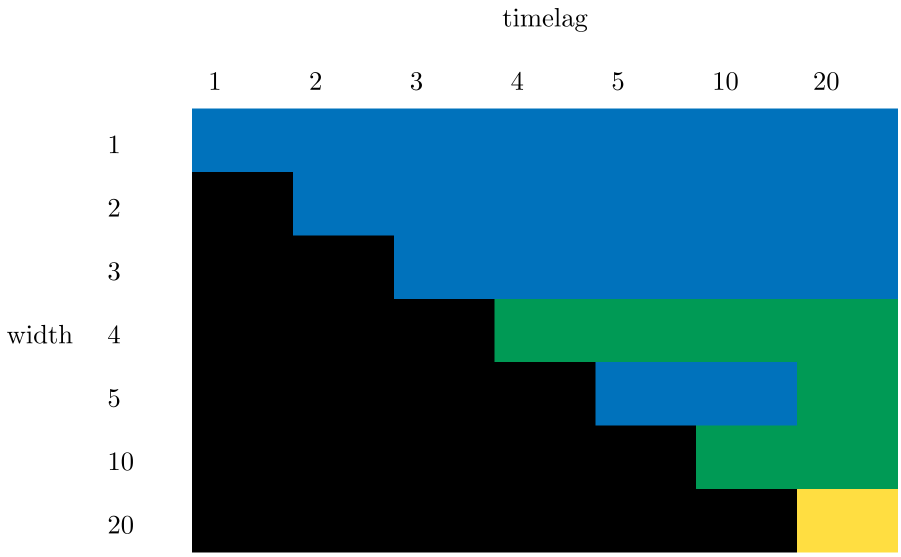
3.1.2 応答関数のパラメータ t_{lag} と w の関係

この章では人工データの t_{lag} と w を変化させて、正しく推定できる限界と 2 つのパラメータの関係性について調べる。t_{lag} は 1、2、3、4、5、10、20 と変化させる。w は 1、2、3、4、5、10、20 と変化させる。ただし、t_{lag} よりも大きい w は応答関数の定義より、因果律に反するためここでは考えない。実験結果を表 3.1 に示す。この表では横軸に t_{lag}、縦軸に w を取り、両方のパラメータの推定に成功しているときは青、t_{lag} のみ成功しており w には真値付近の他に有意な局所解が見られる場合には緑、さらに w の推定にも失敗している時は黄色で示している。表 3.1 中の青の部分の代表例として t_{lag} が 20、w が 3 の事後確率を図 3.4 に、緑の部分の代表例として t_{lag} が 20、w が 5 の事後確率を図 3.5 に、黄色の部分の代表例として t_{lag} が 20、w が 20 の事後確率を図 3.6 に示す。なおこれらの事後確率を得る前に、最初に観測基線長である -100 から 100 までの広い範囲で t_{lag} を推定し、真値付近に大局所解があることを確認して、真値付近に狭い事前分布を設定している。今回の事後確率は t_{lag} に 15 から 25 の事前分布を置いている。表 3.1 から分かるように t_{lag} は全ての組み合わせにおいて推定には成功する。一方で w の推定は 3 以下の時は全ての組み合わせで成功するが、4 以上の時は真値以外にも有意な局所解が見られたり、推定に失敗することが分かる。図 3.4 では、t_{lag}、w において真値付近に分布の中心があることを確認できる。図 3.5 では、t_{lag} には大局解以外にも局所解があるが大局解の確率が大きいので t_{lag} の推定には成功していると言える。しかし、w は真値の 5 の確率よりも 2 付近の確率の方が大きいので推定に失敗していることが確認できる。図 3.6 では、t_{lag} の分布が真値付近に幅広い裾を持っているが分布の中心は真値付近にあるので推定に成功していると言える。しかし、w は真値の 20 の確率がほぼ 0 なので推定には失敗していると言える。

τ が 5 の時は、w が 4 以上なら w の推定に失敗することが分かった。表 3.1 からわかるように、w の推定は一様に $w \geq 4$ で失敗しており、これは推定に失敗する w の値は t_{lag} には依存していないことを示している。

大きな w の時に推定に失敗する原因については 4.1 章で議論する。

表 3.1: 様々な t_{lag} と w での推定の成否。横軸に t_{lag}、縦軸に w を取り、両方のパラメータの推定に成功しているときは青、t_{lag} のみ成功しており w には真値付近の他に有意な局所解が見られる場合には緑、さらに w の推定に失敗している時は黄色で示している。上記の因果律に反する時は黒で示している。



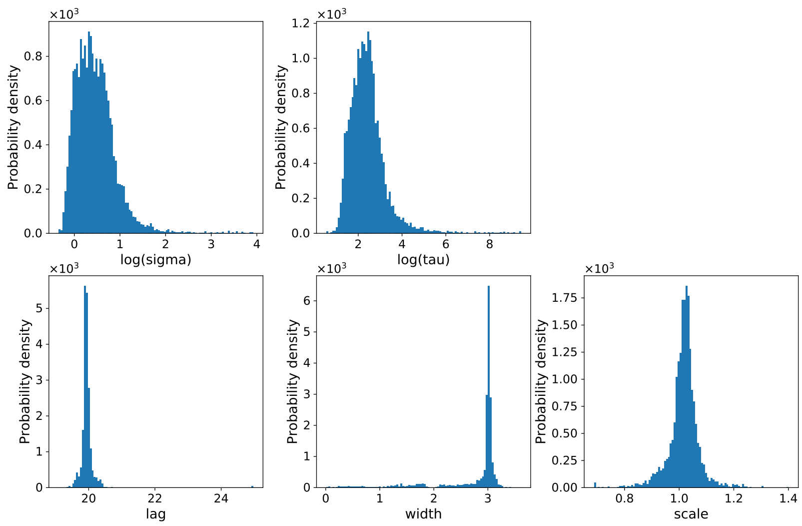


図 3.4: t_{lag} が 20、w が 3 の時の事後確率。t_{lag} には 15 から 25 の事前分布を置いている。横軸は上段左から σ、τ、下段左から t_{lag}、w、A である。縦軸は確率密度である。t_{lag}、w ともに推定に成功している。

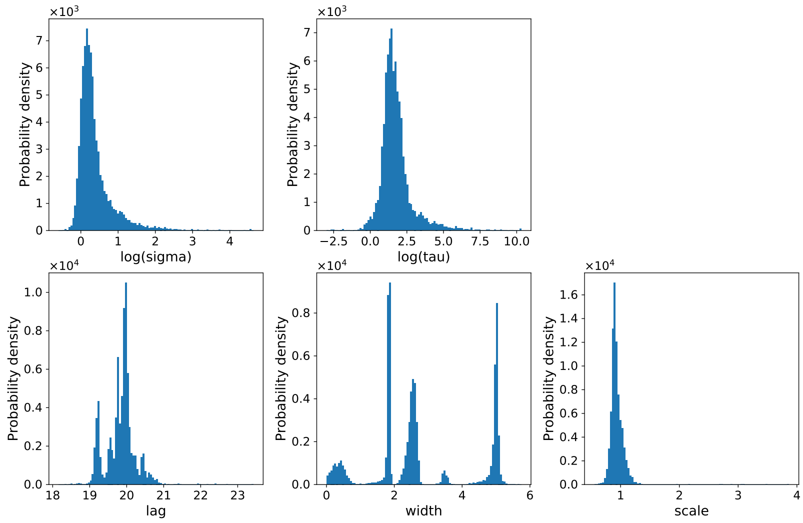


図 3.5: t_{lag} が 20、w が 5 の時の事後確率。t_{lag} には 15 から 25 の事前分布を置いている。横軸は上段左から σ、τ、下段左から t_{lag}、w、A である。縦軸は確率密度である。t_{lag} の推定には成功しているが w は真値以外にも有意な局所解が確認できる。

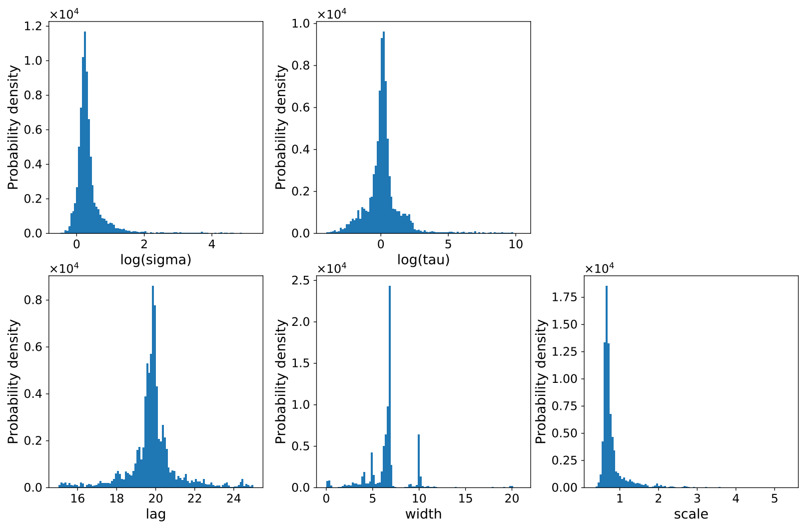


図 3.6: t_{lag} が 20、w が 20 の時の事後確率。t_{lag} には 15 から 25 の事前分布を置いている。横軸は上段左から σ、τ、下段左から t_{lag}、w、A である。縦軸は確率密度である。w が 20 のところでの確率がほぼ 0 に等しく推定に失敗していることが確認できる。

3.1.3　観測間隔よりも小さい $t_{\rm lag}$ の推定

この章では、データ間隔よりも小さい $t_{\rm lag}$ の推定を行う。これらの推定は、相関係数や DCF を用いたタイムラグ推定では不可能なため、OU 過程を用いたモデルで推定ができれば価値は高い。実験は $t_{\rm lag}$ を 0.1 から 0.1 刻みで増加させ、w は 0 で行う。その他のパラメータは既定値で行う。

結果は $t = 0.1$ の時でも推定に成功することが分かった。この時の事後確率を図 3.7 に、推定後の光度曲線を図 3.8 に示す。3.1.2 節と同様に観測基線長に相当する −100–+100 の範囲で $t_{\rm lag}$ を推定し、真値付近に大局所解があることを確認して、狭い範囲に事前分布を設定している。ここでは $t_{\rm lag}$ に 0 から 1 の事前分布を置いている。図 3.7 より確かに $t_{\rm lag}$ の分布の中心が 0.1 付近にあり、w は 0 に分布の中心があることが分かる。また図 3.8 では 0.1 ずれたデータに対して正しく光度曲線モデルで再現されている。

この手法でデータ間隔の少なくとも 10%の $t_{\rm lag}$ を推定できることが分かった。

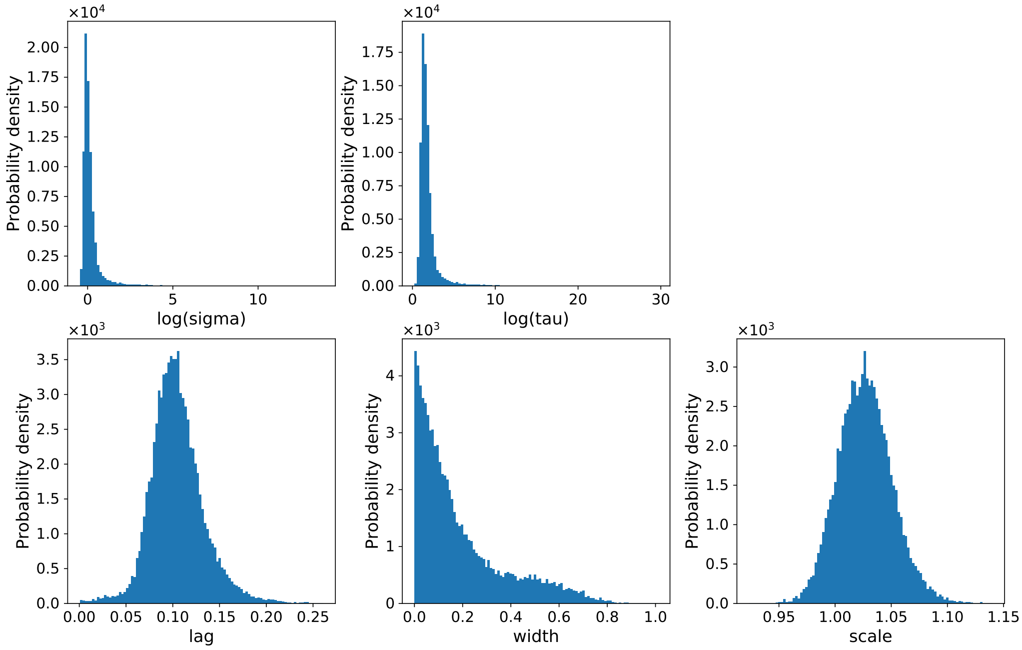


図 3.7: $t_{\rm lag}$ が 0.1、w が 0 の時の事後確率。$t_{\rm lag}$ に 0 から 1 の事前分布を置いている。横軸は上段左から σ、τ、下段左から $t_{\rm lag}$、w、A である。縦軸は確率密度である。

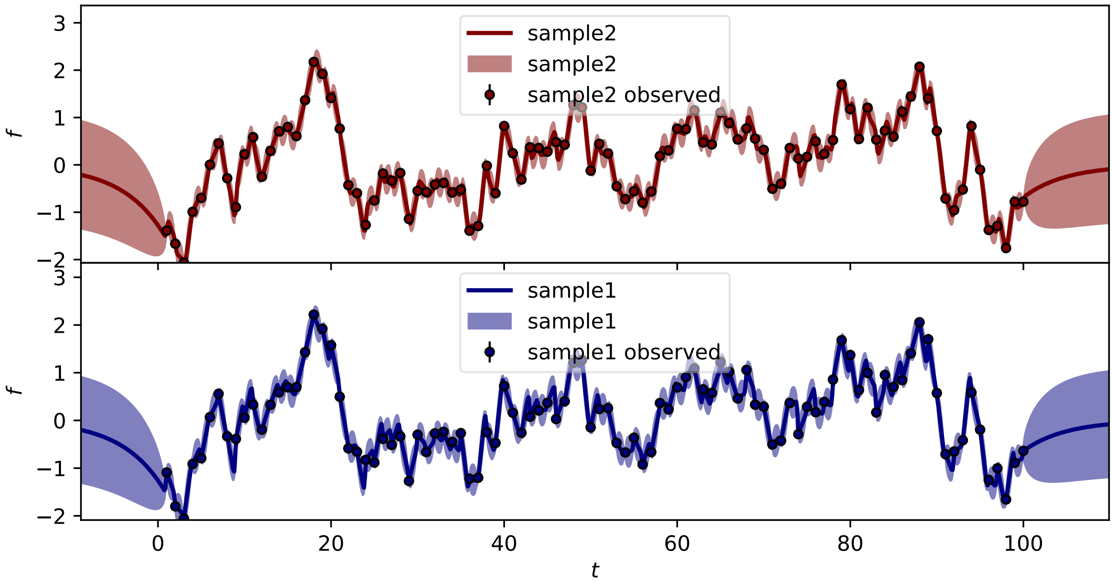


図 3.8: t_{lag} が 0.1、w が 0 の時の光度曲線。横軸が時間、縦軸が光度を示している。

3.1.4 欠損データ

現代の天文学においてデータが長期間に渡って一定の間隔で得られることは稀であり、特に地上観測では季節や天候の影響でデータの欠損が生じることが多い。データの欠損はランダムに生じることもあるが一定期間生じることもある。このような欠損のパターンによって推定の成否に影響を与える可能性があるので今回の実験では、ランダムに欠損がある時と一定間隔欠損がある時の 2 つの場合について実験を行う。この章では、パラメータ推定に成功する限界のデータ欠損について調べる。パラメータは全て既定値で行い、完全にデータがある時の数を 100 とし、そこからデータを減らしていく。

初めにランダムに欠損がある場合の推定の成否を調べる。以下、OU 過程に従うサンプルを生成し、正規乱数の誤差を足して生成したサンプルをサンプル 1 と呼び、t_{lag}、w、A で定義される応答関数によってサンプル 1 から生成されたサンプルをサンプル 2 と呼ぶ。まず、サンプル 1 のデータのみを用いて、σ、τ の推定に成功するために必要なデータ数を調べる。次に、サンプル 1 は完全にある、つまり 100 個データが得られていて、サンプル 2 に欠損がある時の実験を行う。最後に、両方のサンプルに欠損がある時の実験を行う。この時簡単のために両方のサンプルに同じ数の欠損がある時を考える。

次に一定間隔で欠損がある時を実験を行う。ランダムに欠損がある時と同様に、サンプル 1 のみの推定、サンプル 1 はデータが 100 個得られサンプル 2 に欠損がある時、両方のサンプルに欠損がある時の実験を行う。

ランダムなデータ欠損がある時系列データの OU 過程パラメータの推定

この章では、サンプル 1 にデータ欠損がある時の σ と τ の推定を行う。1 つのサンプルに対して σ と τ の推定を行うことにより、2 つのサンプルの推定に成功する必要最低限のデータ数を見積もることを目的と

する。片方のサンプルのみに欠損がある時や両方のサンプルがある時の実験では、本実験で見積もられたデータ数よりも少ないデータ数での推定は失敗することが予測できる。

3.1.1 章で述べたように、これらのパラメータはほとんどが 95%信頼区間に入るもの分布全体の幅が広いことが多く、信頼区間内に入るかどうかで成否を判断するのは難しい。しかし、データ数を増やしていくと信頼区間の幅が狭くなっていく傾向があるので、95%信頼区間の幅で推定の評価を行うことにした。また、σ と τ の 2 つのパラメータがあるが、本実験では光度変動の激しさよりも、時間的な変動スケールに注目した実験なので、τ を推定で着目するパラメータとする。

サンプル 1 にランダムな欠損を与えた時の、データ数と τ の 95%信頼区間の幅の相関を図 3.9 に示す。図 3.9 より、データ点が数個しかない時は、τ の分布の幅が観測基線長（$= 100$）よりも長くなってしまうことがわかる。観測基線長よりも幅が長い信頼区間は、実用上では意味がないので推定に失敗していると判断する。データが約 20 点以上得られると、サンプルによっては信頼区間の幅が 40 ぐらいになることはあるものの、ほとんどが 20 以下であることが分かり、さらに、20 点以上データを増やしても推定の精度がよくなることはないことも分かる。3.1.1 章で述べたことも踏まえて、サンプルが 1 つの時は、全体のデータ数が 100 点に対して約 20 点以上得られていれば十分推定に成功すると考える。

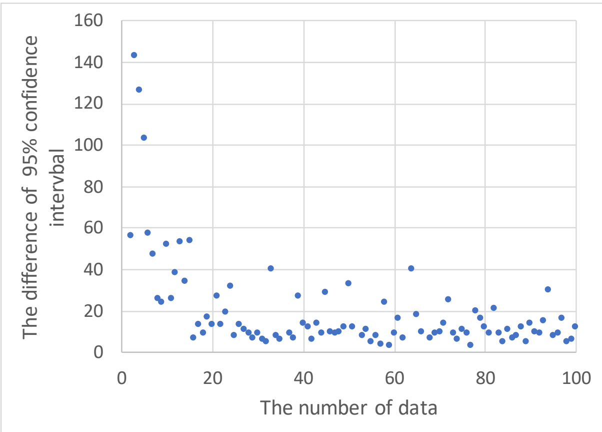


図 3.9: ランダムな欠損がある時の、データ数と τ の 95%信頼区間の幅の相関。横軸がデータ数、縦軸が 95%信頼区間の幅である。

完全なデータとランダムなデータ欠損がある場合のタイムラグ推定

次にサンプル 1 のデータが 100 個得られ、サンプル 2 にランダムな欠損がある時の実験を行う。

まず始めに、サンプル 2 が最も少ないデータ数で推定に成功した例としてサンプル 2 のデータ数が 15 点の時の事前分布なしの事後確率を図 3.10 に、t_{lag} に 15 から 21 の事前分布を置いた時の事後確率を図 3.11 に、推定後の光度曲線の図を 3.12 に示す。事前分布なしの事後確率の図 3.10 より、データが少ないと t_{lag} の真値（$t = 20$）付近の他にも有意な局所解が見られるが、$t \sim \pm 100$ 付近の解は観測期間に匹敵するほど長いので現実的な解ではない。また、現実的な範囲では真値付近の解が最も確率が高い。従って、この真値付近の解周辺に t_{lag} の事前分布を仮定する判断は妥当なものだと考えられる。なお、t_{lag} が 20 付近に事前分布が見られた時は、基本的に 15 から 25 に事前分布を仮定する。しかし、20 付近の大局解の他に有意な局所解が見られた時にはその解を含まない範囲に事前分布を仮定する。今回は有意な局所解が見られたの

で 15 から 21 に事前分布を置いた。t_{lag} に 15 から 21 の事前分布を置いた時の事後確率の図 3.11 では、w が 2 付近に分布の中心があり、事前分布なしの時よりも真値を推定できていることがわかる。図 3.12 では、サンプル 2 のデータが 15 点しかない場合でも、モデルが高い精度で観測を再現できていることがわかる。また、サンプル 2 において時間が 0 から 20 付近までは信頼区間の広がりが大きいことが確認できる。このことはこれらと相関があると考えられるサンプル 1 のデータが存在しないために、欠損部分の信頼区間を小さくすることが難しいことを示している。

次に、t_{lag} と w の推定に失敗した例として、サンプル 2 が 2 点の時の事前分布なしの事後確率を図 3.13 に示す。図 3.13 をみると t_{lag} には意味のある解が見られず推定に失敗していることが分かる。

このようにデータ数が多いと推定に成功しやすく、データ数が少ないと推定に失敗しやすいことが分かる。データ数を変えた時の t_{lag} と w の推定の結果を図 3.14 に示す。これは横軸にデータ数、縦軸に推定の成否をとったグラフである。図中には t_{lag} と w の成否が描かれている。図 3.14 から分かるように t_{lag} の推定はサンプル 2 のデータ数が 4 点以上あると推定に成功することが分かる。w の推定は 20 点から 30 点ほどデータが得られると成功し始め、30 点以上データが得られると失敗しなくなることが分かる。

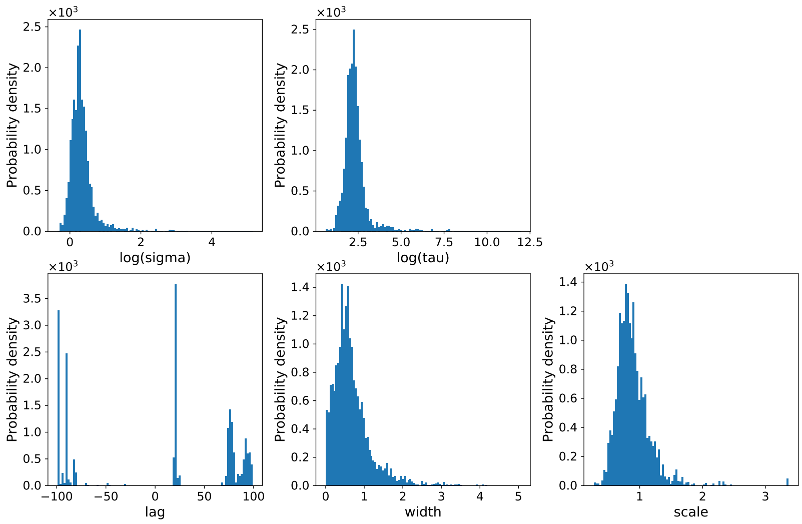


図 3.10: t_{lag} が 20、w が 2、サンプル 1 が 100 点、サンプル 2 が 15 点で、事前分布なしの事後確率。

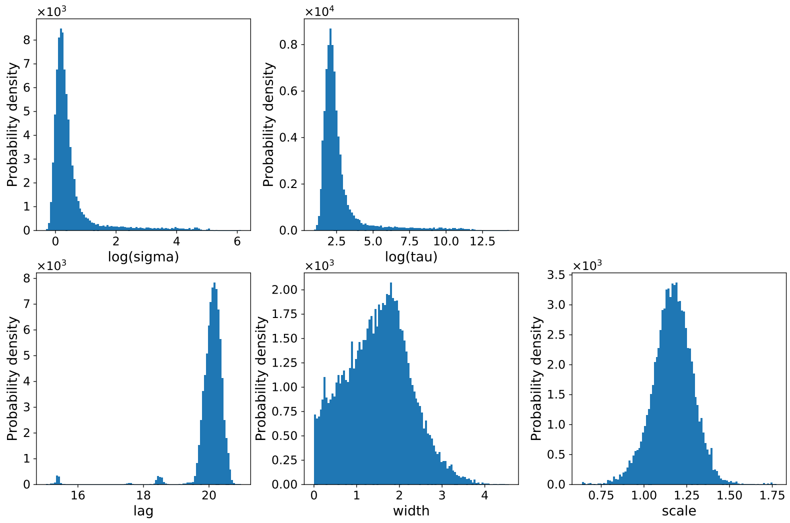


図 3.11: t_{lag} が 20、w が 2、サンプル 1 が 100 点、サンプル 2 が 15 点で、t_{lag} に 15 から 25 の事前分布を置いた時の事後確率。

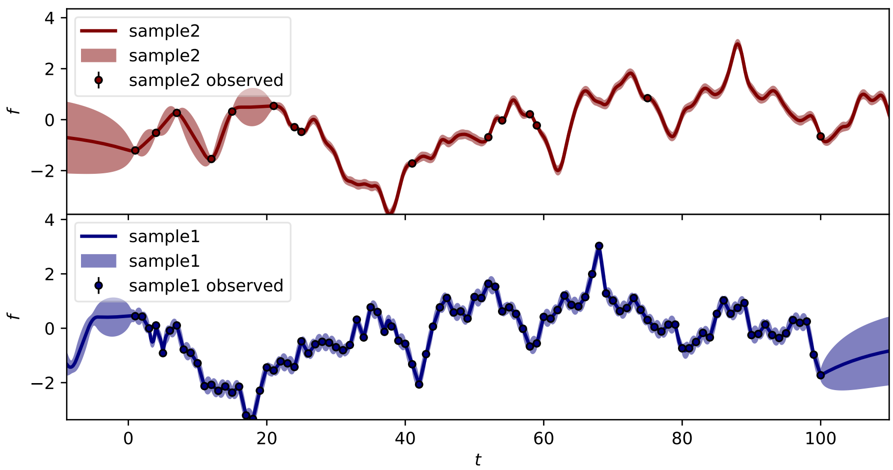


図 3.12: t_{lag} が 20、w が 2、サンプル 1 が 100 点、サンプル 2 にランダム欠損を与えデータが 15 点ある時の光度曲線。

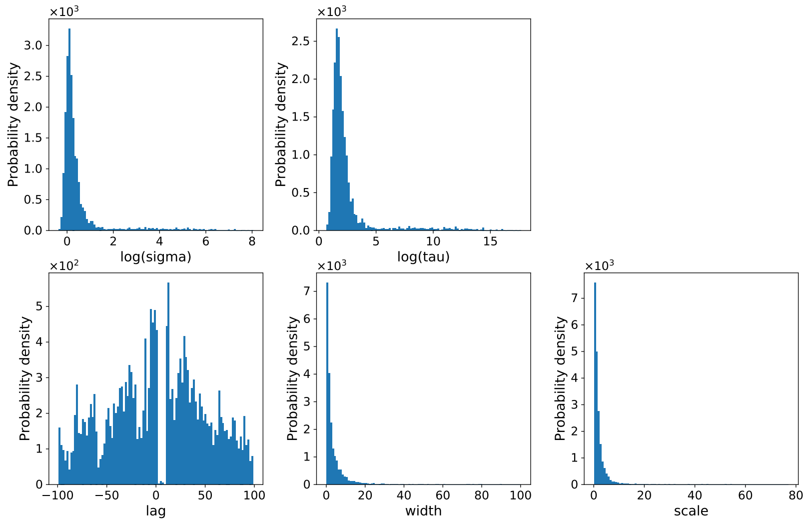


図 3.13: t_{lag} が 20、w が 2、サンプル 1 が 100 点、サンプル 2 が 3 点で、事前分布なしの事後確率。

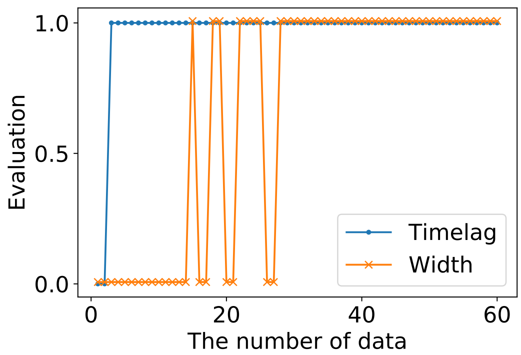


図 3.14: サンプル 1 は 100 個データが得られ、サンプル 2 にランダムに欠損がある時の、t_{lag} と w の推定の成否。0 が推定に失敗、1 が推定に成功。

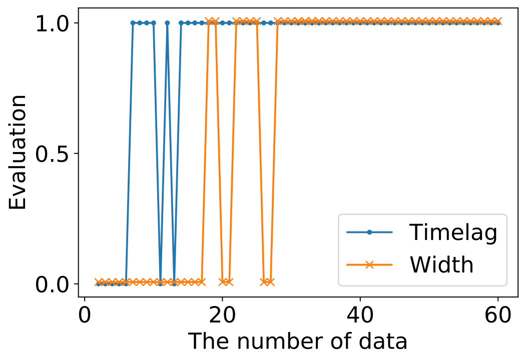


図 3.15: サンプル 1 とサンプル 2 に同じ数のランダムな欠損がある時の、t_{lag} と w の推定の成否。0 が推定に失敗、1 が推定に成功。

両方のサンプルにランダムなデータ欠損がある場合のタイムラグ推定

最後にサンプル 1 とサンプル 2 に同じ数のランダムな欠損があった時の結果を図 3.15 に示す。この時の最も少ないデータ数で推定に成功した例として両方のサンプルにデータが 18 点ある時の例を示すのが、上述のように、この場合でも最初に t_{lag} を広い範囲で推定し、真値付近に最も確率が高い解があることを確認し、次に真値付近の狭い範囲で t_{lag} を推定した。以下の実験でも同様の手続きで解析しており、t_{lag} を広い範囲で推定した結果の事後確率は以降では省略する。両方のデータが 18 点あり、t_{lag} に 15 から 25 までの

事前分布を置いた時の事後確率を図 3.16 に示す。事前分布を置いた時の光度曲線を図 3.17 に示す。図 3.15 からわかるように、t_{lag} はデータ数が 7 点ぐらい得られると推定に成功してくることが分かり、データ数が 12 以上になると推定に失敗しなくなることが分かる。w はデータ数が 18 点ぐらいから推定に成功し始め、30 点以上得られると失敗しなくなることが分かる。t_{lag} に 15 から 25 の事前分布を置いた時の事後確率の図 3.16 では、t_{lag} と w は, どちらも真値である 20 と 2 付近に分布の中心があることが確認できる。光度曲線の図 3.17 では、これぐらい欠損のある時系列データでは一見して相関があるようには見えないが、両方のデータが共通しての OU 過程に厳密に従う状況においては、t_{lag} と w の推定が可能であることがわかる。

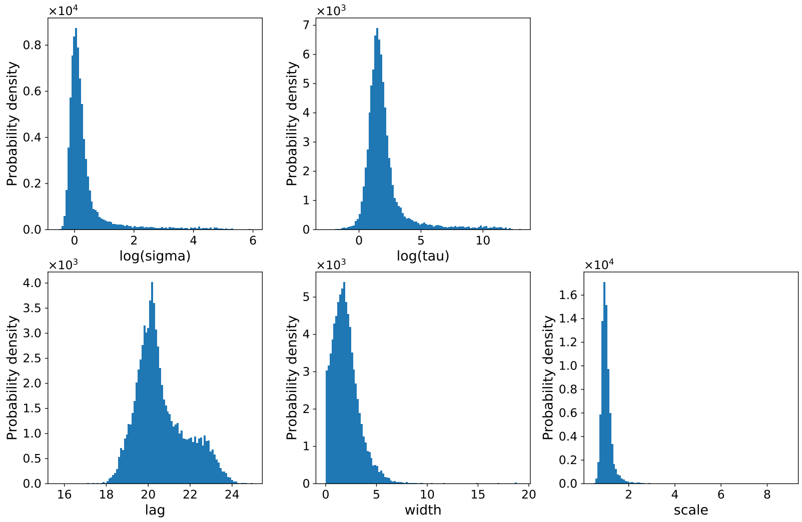


図 3.16: t_{lag} が 20、w が 2、サンプル 1 とサンプル 2 が共に 18 点で、t_{lag} に 15 から 25 の事前分布を置いた時の事後確率。

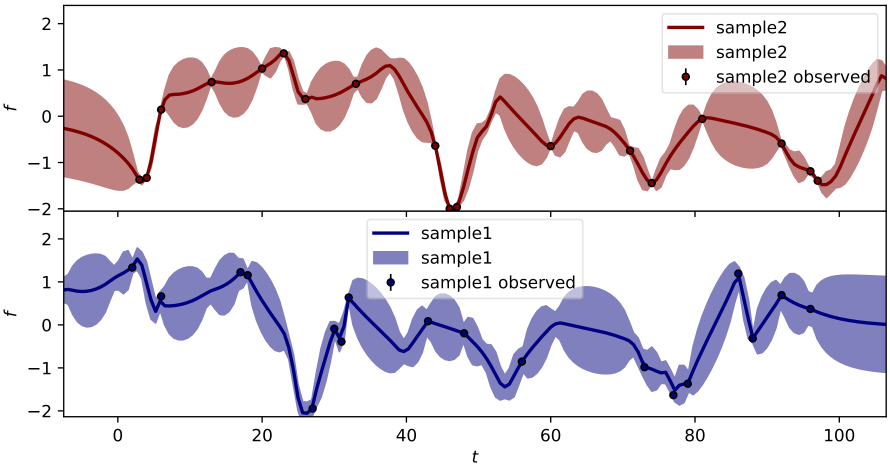


図 3.17: $t_{\rm lag}$ が 20、w が 2、サンプル 1 とサンプル 2 が共に 18 点で、$t_{\rm lag}$ に 15 から 25 の事前分布を置いた時の光度曲線。

一定間隔でデータ欠損がある時系列データの OU 過程パラメータの推定

次に、ランダムな欠損ではなく、一定の期間に欠損がある場合の推定の成否について調べる。本実験では、データの欠損が生じている間隔をデータ 18 点で固定し、この間隔の数を増やすことで、データ数を 10、28、46、64、82 と変化させる。なお、欠損部分が繋がることで欠損期間が実質的に長くなり、観測期間が 100 より短くなることを防ぐ目的で欠損間隔の両端のデータを残している。これにより連続した欠損部分の期間が一定になるようにしている。そのため最小のデータ数は 10 点となる。

3.1.4 章と同様の理由から τ の信頼区間の幅で推定の評価を行うことにする。一定間隔の欠損がある時では前述したように、データ数が 10、28、46、64、82 と 5 つのパターンしかデータを得ることができなく、ランダムに欠損を与えた時よりも得られるデータ数が少なくなってしまう。そこで本実験では、データ数が 10 の実験を 10 回、28 の時の実験を 10 回ーのように、同じ欠損のパターンを 10 回行うことにする。

サンプル 1 に一定間隔な欠損を与えた時の、データ数と τ の 95%信頼区間の幅の相関を図 3.18 に示す。図 3.18 より、データが 10 点の時は信頼区間の幅が観測基線長となることもあり、推定に失敗しやすいことが分かる。3.1.4 章の時と同様に、データが 28 点以上データが得られている時は、データ数が多くなっても精度がよくなることはない。よって、サンプルが 1 つの時には少なくとも 28 点以上データがあれば τ は制約できることが分かる。ランダムな欠損を考えた時は 30 点以上データがあれば推定に成功したため、ランダムな欠損と一定間隔の欠損で τ の推定にはほぼ差がないことがわかった。

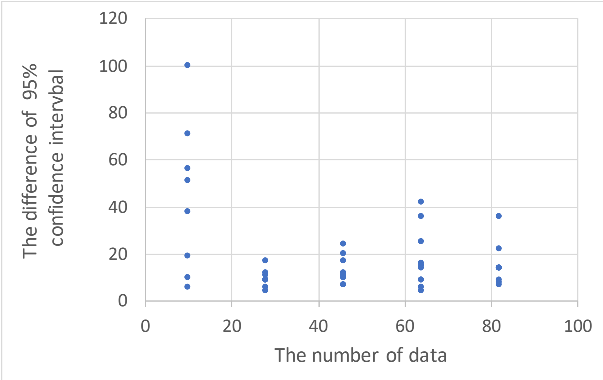


図 3.18: 一定間隔の欠損がある時の、データ数と τ の 95%信頼区間の幅の相関。横軸がデータ数、縦軸が 95%信頼区間の幅である。

完全なデータと一定間隔でデータ欠損がある場合のタイムラグ推定

次にサンプル 1 が 100 点得られ、サンプル 2 に一定間隔の欠損がある時の実験を行う。上述したようにサンプル 2 のデータ数は 10、28、46、64、82 と変化させる。

実験の結果、最小のデータ数である 10 点の時も成功することが分かった。この時、$t_{\rm lag}$ に 15 から 22 の事前分布を置いた事後確率を図 3.19 に、事前分布を置いた時の光度曲線を図 3.20 に示す。$t_{\rm lag}$ に 15 から 22 に事前分布を置いた事後確率の図 3.19 では、$t_{\rm lag}$ は 20 付近、w は 2 付近に分布の中心があることが確認できる。この時の光度曲線の図 3.20 を見ると、0 から 20 の疎な期間では信頼区間が大きいことが確認できる。これは、この部分と相関があると思われるサンプル 1 のデータがないために信頼区間を小さくすることができないからである。

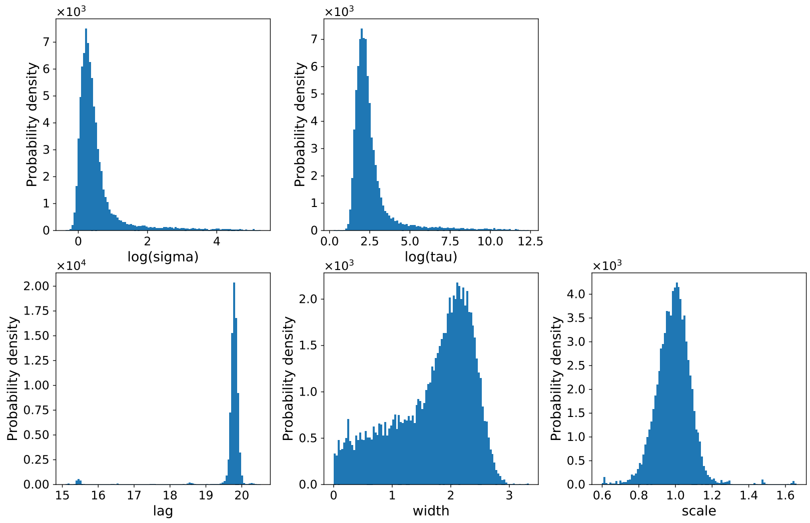


図 3.19: t_{lag} が 20、w が 2、サンプル 1 が 100 点、サンプル 2 が 10 点で、t_{lag} に 15 から 22 の事前分布を置いた時の事後確率。

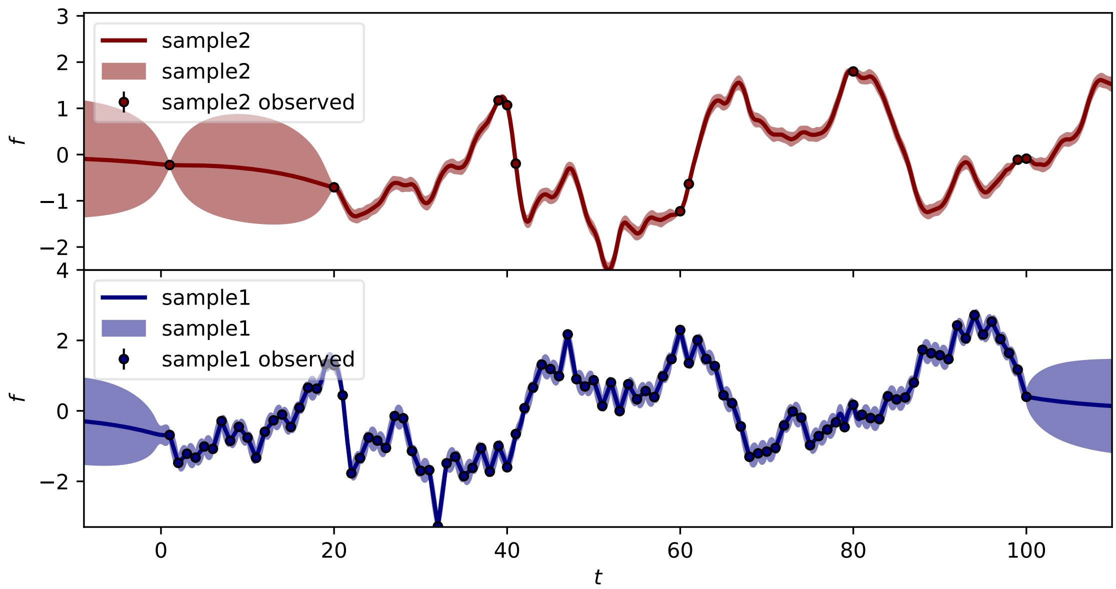


図 3.20: t_{lag} が 20、w が 2、サンプル 1 が 100 点、サンプル 2 に一定間隔の欠損があり、データ数が 10 点ある時の光度曲線。

両方のサンプルに一定間隔でデータ欠損がある場合のタイムラグ推定

最後にサンプル 1 とサンプル 2 に同じ数のデータ欠損がある時の結果を図 3.21 に示す。今回推定に成功した最小の例として共にデータ数が 28 点あり、t_{lag} に 15 から 25 までの事前分布を置いた事後確率を図 3.22 に示す。事前分布を置いた時の光度曲線を図 3.23 に示す。図 3.21 より、共にデータ数が 10 点得られている時は t_{lag} の推定には成功するが、w の推定には失敗することが分かる。しかしお互いに 28 個以上のデータ、つまり合わせて 56 個以上データが得られている時は t_{lag}、w 共に推定に成功することが分かる。t_{lag} に 15 から 25 の事前分布を置いた時の事後確率の図 3.22 より、確かに t_{lag} と w の分布の中心が真値の 20 と 2 付近にあることが確認できる。この時の光度曲線の図 3.23 では、0 から 20 付近の密な期間は、信頼区間が小さく精度よく推定できているが、それ以外の疎な期間は信頼区間が大きく推定の精度が悪いことが確認できる。

以上の実験の結果、データ数が少ない場合でも t_{lag} の推定に成功することが分かった。しかし、本実験は測定誤差が無視できるほど小さい場合を扱っており誤差が無視できない場合に同様の推定に成功するかは自明ではない。さらに、2 つのデータ間の相関が自明ではないほどにデータ数が少ない場合でも推定に成功することは、本来相関していないデータについても偽の相関が推定される可能性を示唆する。これらのことは 4.4.1 章と 4.4.2 章で議論する。

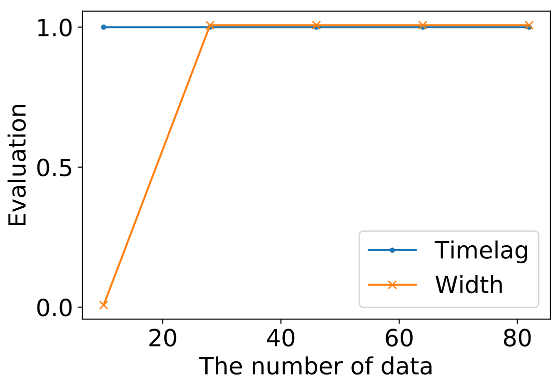


図 3.21: サンプル 1 とサンプル 2 に同じ数の一定間隔の欠損がある時の、t_{lag} と w の推定の成否。0 が推定に失敗、1 が推定に成功。

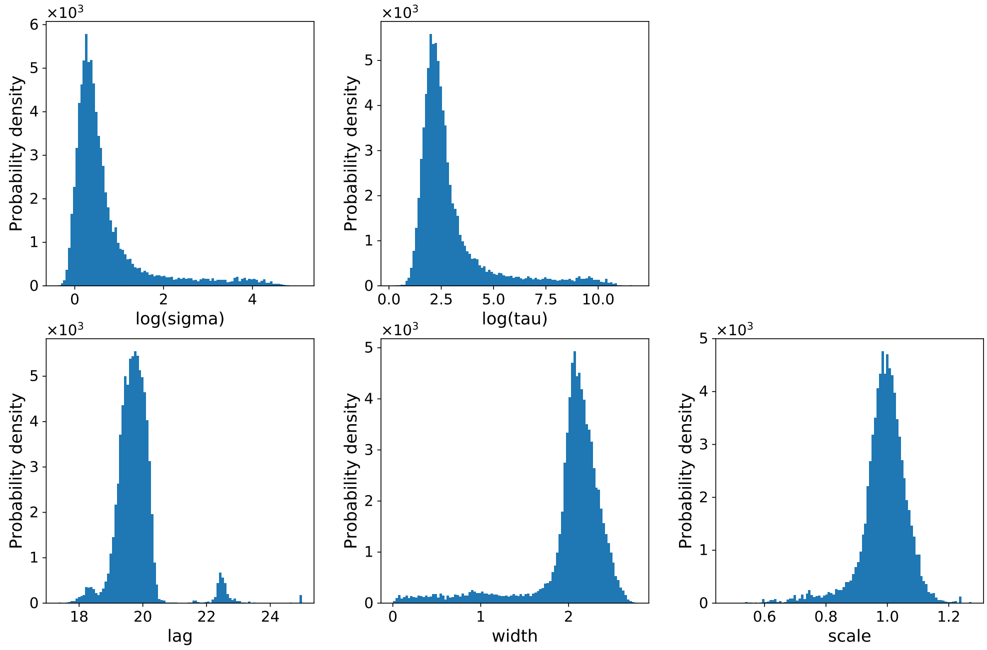


図 3.22: t_{lag} が 20、w が 2、サンプル 1 とサンプル 2 のデータ数が共に 28 点で、t_{lag} に 15 から 25 の事前分布を置いた時の事後確率。

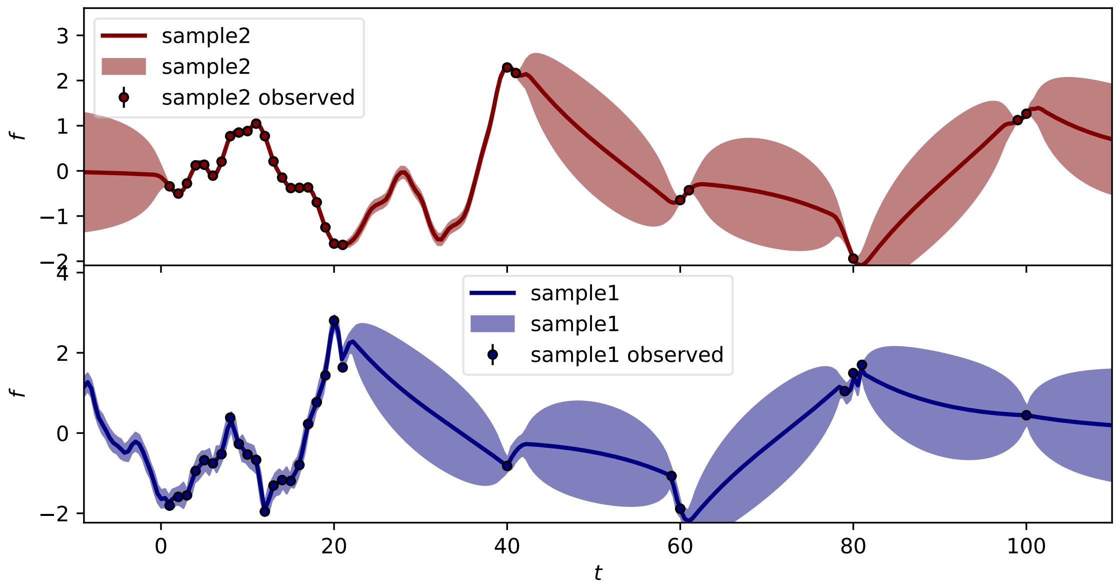


図 3.23: t_{lag} が 20、w が 2、サンプル 1 とサンプル 2 が共に 28 点ある時の光度曲線。

3.1.5　データの誤差

これまでの実験ではデータの測定誤差は無視できるほど小さいと仮定してきたが、実際のデータには無視できない測定誤差が含まれることが多い。ここでは測定誤差の大きさが推定の成否に与える影響を調べる。

実験で用いるパラメータは既定値である。データの誤差は正規分布に従い、その標準偏差を誤差とする。図中で表示するエラーバーは、標準偏差と同じ大きさで与える。

データに誤差を与えた時の実験結果を図 3.24 に示す。今回、最も大きい誤差で推定に成功した例として、t_{lag} に 15 から 25 の事前分布を置いた事後確率を図 3.25 に、事前分布を置いた時の光度曲線を図 3.26 に示す。図 3.24 より、t_{lag} は誤差の標準偏差が 1.5 までは推定に成功し、それ以降は推定に失敗しやすくなることが分かる。ここで、本実験ではデータは典型的に $\mathcal{O}(1)$ の値をもつため、例えば標準偏差 1.5 の誤差は観測値の 150%誤差とみなすことができる。

また w は誤差が 1.0 付近まで推定に成功し、それ以降は推定に失敗しやすくなることがわかる。事前分布を置いた事後確率の図 3.25 では t_{lag}、w 共に、真値付近に分布の中心があることを確認できる。

事前分布を置いた時の光度曲線の図 3.26 では光度の時間変動に対してエラーがかなり大きいことが分かるが、データに欠損がなく完全で、両データともに OU 過程に従うことが分かっている場合は、このように推定に成功することが分かった。

次に誤差が大きすぎて t_{lag} の推定に失敗した例として、誤差が 1.6 の時の事後確率を図 3.27 に示す。図 3.27 より、t_{lag} が真値の 20 の確率が大きくなっているが観測基線長に匹敵するほどの大きな裾を持っているので推定に失敗していると言える。

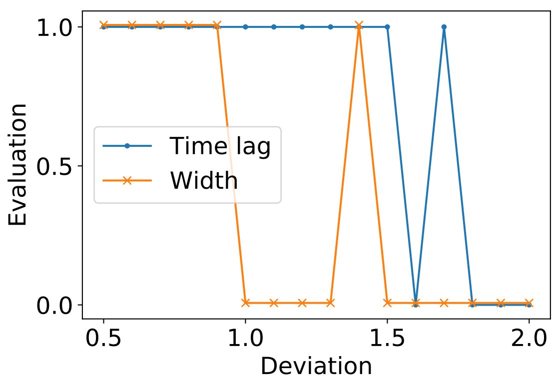


図 3.24: データに誤差を与えた時の t_{lag} と w の推定の成否。縦軸は、0 が推定に失敗、1 が推定に成功。横軸はデータに与えた誤差である。

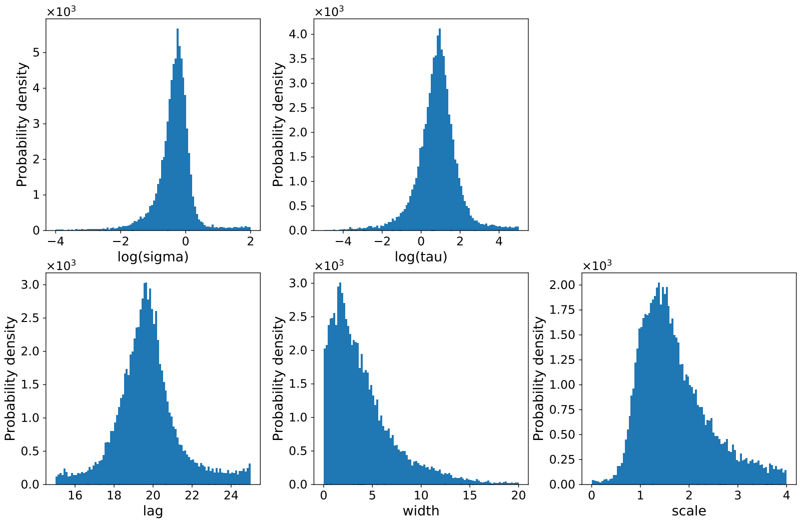


図 3.25: 誤差の標準偏差が 1.4 で、t_{lag} に 15 から 25 の事前分布を置いた時の事後確率。横軸は上段左から σ、τ、下段左から t_{lag}、w、A である。縦軸は確率密度である。

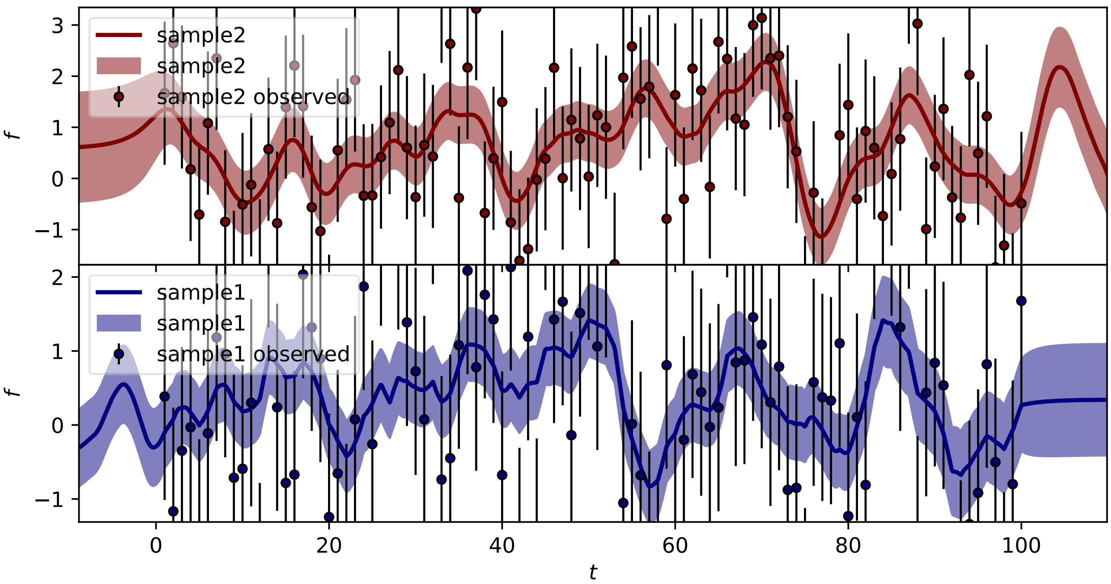


図 3.26: 誤差の標準偏差が 1.4 の時の光度曲線。

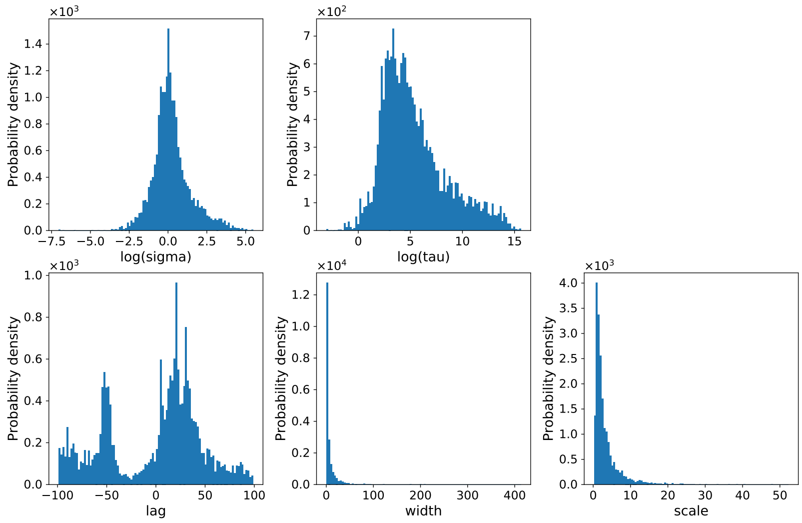


図 3.27: 誤差の標準偏差が 1.6 の時の事後確率。横軸は上段左から σ、τ、下段左から t_{lag}、w、A である。縦軸は確率密度である。

3.2 活動銀河核 PKS 1510−089

次に、実際のデータへの応用例として、活動銀河核 PKS 1510−089 のデータを用いたタイムラグ等のパラメータ推定を行う。この天体は先行研究でタイムラグがあらかじめ分かっている。この天体は、先行研究でガンマ線と電波の光度曲線間のタイムラグが、フレアの極大時刻の差から 54 日と結論づけている (Beaklini et al., 2017)。天体のガンマ線と電波の光度曲線を図 3.28 に示す。縦軸が光度、横軸がユリウス通日（JD）である。光度は自然対数をとり、平均で引いている。電波は Beaklini et al. (2017) に記載されているものである。データはブラジルの Itapetinga Radio Observartory の 13.7 m 電波望遠鏡で観測されたもので、波長は 7 mm である。ガンマ線はフェルミガンマ線宇宙望遠鏡で観測された公開データで、100 MeV 以上の光子の 1 日ごとのデータである。

図 3.28 を見ると、確かに 50 日ぐらいの t_{lag} があるように見える。また、ガンマ線が鋭いピークなのに対して、電波はピークが鈍っているようにも見える。これが w の効果であると考えることができる。先行研究では t_{lag} しか推定できておらず、w に有意な解があればジェットの研究に新しい情報を加えて議論することができ有益である。また、ガンマ線は観測期間において長期的な変動成分は見られないが、電波の変動は平均的に緩やかに上昇して下降している傾向が見られる。そこでこの章では、最初にどちらもそのままの光度曲線の解析を行い、次に、電波の光度曲線に見られる長期的な変動成分の影響を除去するために低周波成分を引いた光度曲線とそのままのガンマ線の光度曲線で解析を行う。

以下、推定を行っていく。始めに時間軸に変更を加えず推定を行ったところ、t_{lag} は先行研究と同様に 50 日付近に局所解が見られたが、w は 0 日に分布の中心があった。次に、時間軸を 0.1 倍して推定を行ったところ、t_{lag} は同様に 50 日付近に局所解があり、w にも 0 日に大局解があるものの、20 日付近に有意な局所解が見られた。以上の理由から、以下の実験では日にちを 0.1 倍した時の実験結果を示している。

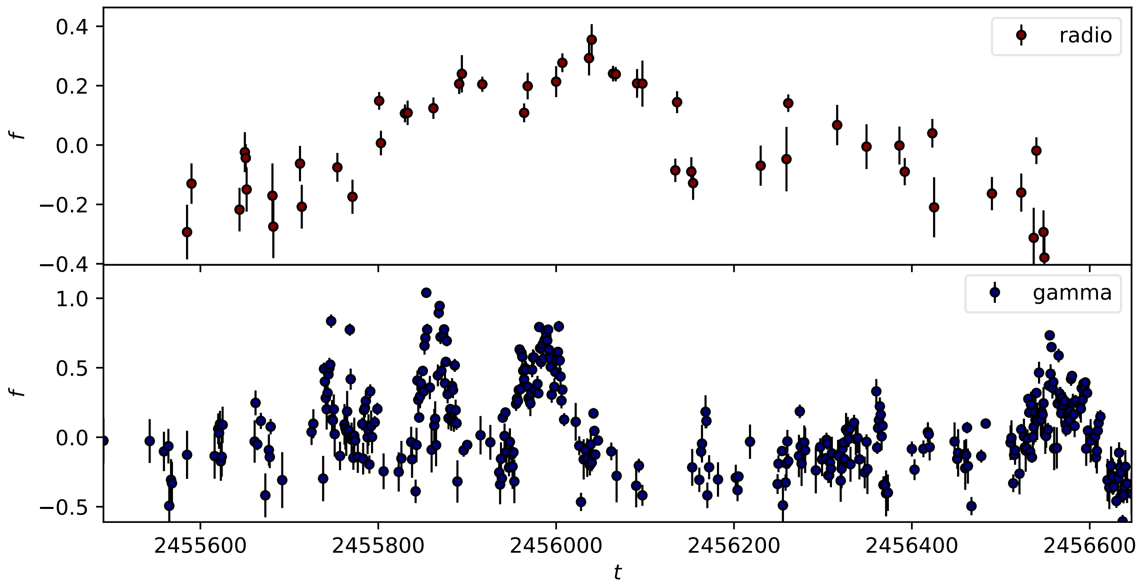


図 3.28: 活動銀河核 PKS 1510−089 の光度曲線。縦軸が光度、横軸が JD である。上が電波、下がガンマ線である。

3.2.1 低周波成分補正がない時の解析

$t_{\rm lag}$ に −100 日から 100 日の事前分布を置いた事後確率を確認し、先行研究で示された 50 日 付近に $t_{\rm lag}$ があることを確認する。この時の事後確率を図 3.29 に示す。図 3.29 より、確かに先行研究のように 50 日付近に有意な解が見られる。そこで $t_{\rm lag}$ に 40 日から 60 日に事前分布を置いた時の事後確率を図 3.30 に示す。図 3.30 より、$t_{\rm lag}$ の局所解が 3 つあることがわかる。それぞれの局所解と w の 0 日付近の局所解と 20 日付近の局所解が対応しているかどうかを確かめるためにそれぞれの局所解付近に事前分布を置いてみたところ、$t_{\rm lag}$ が 46 日付近と 51 日付近の局所解は w が 0 日付近と 20 日付近に対応していること、$t_{\rm lag}$ が 41 日付近の局所解は、w が 0 日付近の局所解しか対応していないことが分かった。しかし、いずれも w が 0 日付近に大局解があることには変わりがなかった。

図 3.30 において最も $t_{\rm lag}$ の確率が大きかった 46 日付近に事前分布をおいた時の事後確率を図 3.31 に、光度曲線を図 3.32 に示す。光度曲線の横軸は JD の 0.1 倍である。図 3.31 に見られる 2 つの w の解の確率を、確率密度を積分することで計算すると、w が 0 日付近の確率が 50%、20 日付近の確率が 30%と求められた。図 3.32 では、ガンマ線の光度曲線に $t_{\rm lag}$ が 46 日、w が 0 日与えた光度曲線が電波のデータに表されていることが分かる。

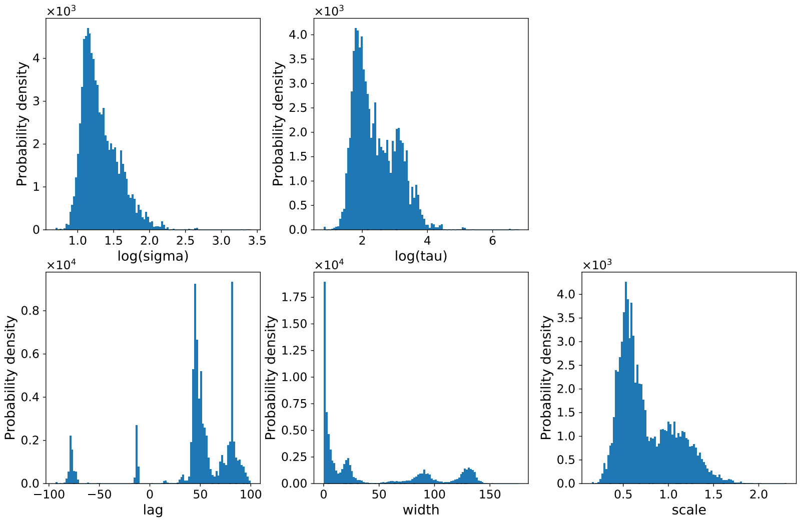


図 3.29: 活動銀河核 PKS 1510−089 の t_{lag} に -100 日から 100 日の事前分布を置いた時の事後確率。横軸は上段左から σ、τ、下段左から t_{lag}、w、A である。縦軸は確率密度である。

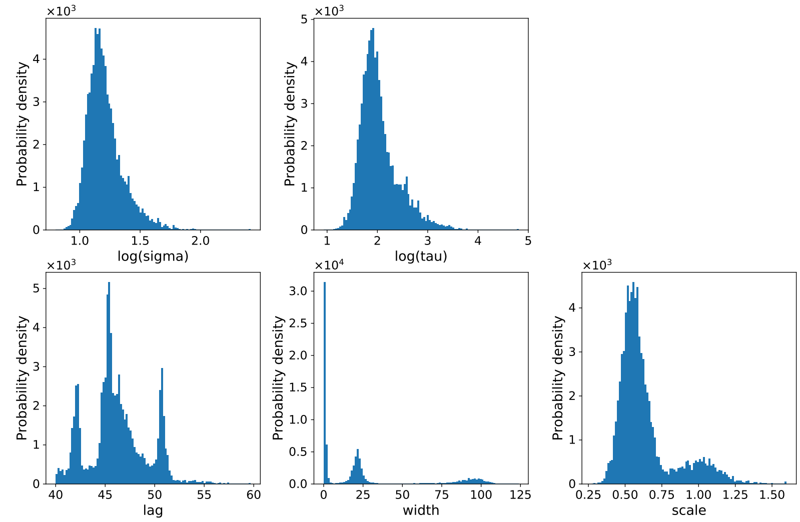


図 3.30: 活動銀河核 PKS 1510−089 の t_{lag} に 40 日から 60 日の事前分布を置いた時の事後確率。横軸は上段左から σ、τ、下段左から t_{lag}、w、A である。縦軸は確率密度である。

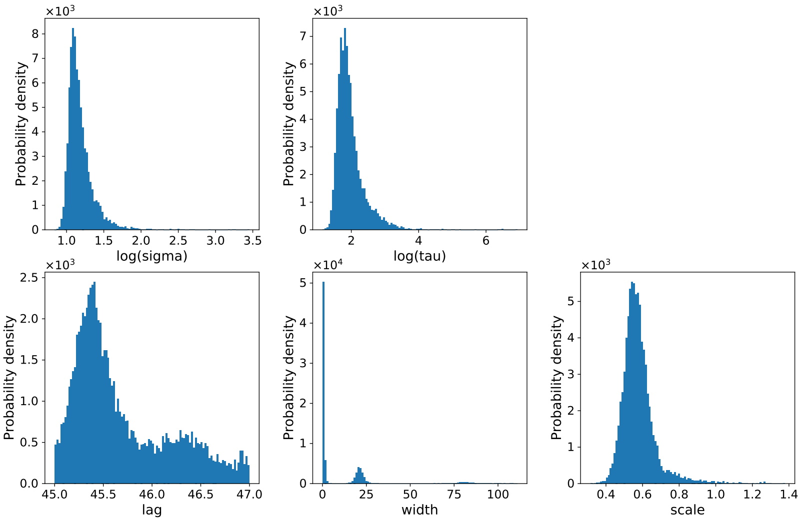


図 3.31: 活動銀河核 PKS 1510−089 の t_{lag} に 45 日から 47 日の事前分布を置いた時の事後確率。横軸は上段左から σ、τ、下段左から t_{lag}、w、A である。縦軸は確率密度である。

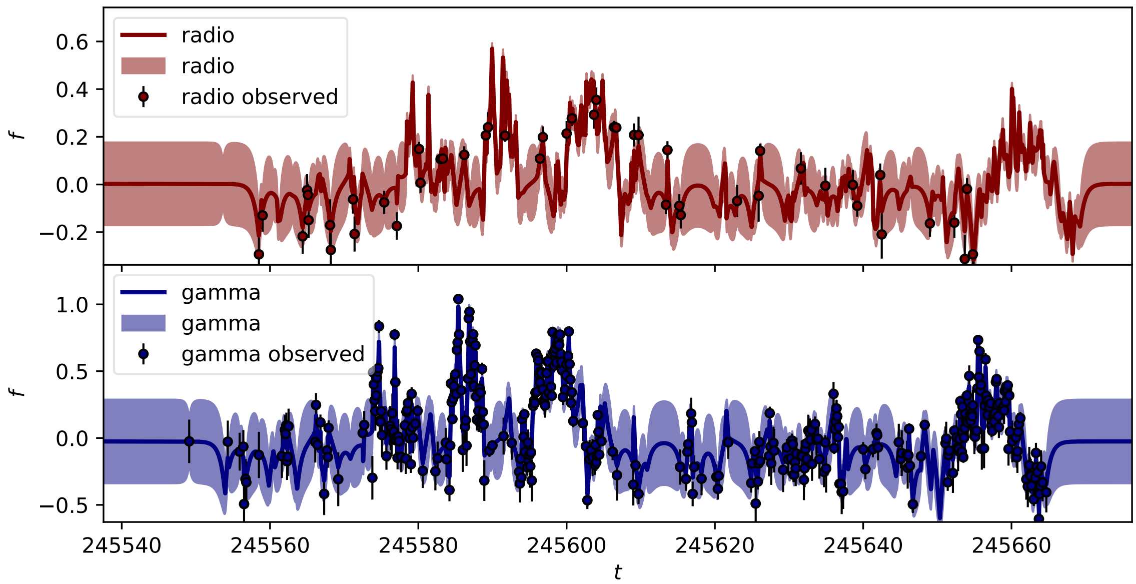


図 3.32: 活動銀河核 PKS 1510−089 の t_{lag} に 45 日から 47 日の事前分布を置いた時の光度曲線。横軸は JD、縦軸は光度の自然対数表記である。上が電波、下がガンマ線の光度曲線である。

3.2.2 低周波成分補正がある時の解析

次に電波に対して低周波成分を引いた光度曲線で推定を行う。図 3.28 をみると、2 次関数的に上昇し、下降していることが確認できるので、2 次回帰曲線を引くことにする。

この光度曲線に対して t_{lag} に 40 日から 60 日の事前分布をおいた時の事後確率を図 3.33 に示す。図 3.33 では図 3.30 と同じように、$t_{\mathrm{lag}} = 42$ 日、46 日、51 日付近に局所解があることが確認でき、$w = 0$ 付近の確率が 60%、$w = 20$ 日付近の確率は 35%であることがわかる。低周波成分を引いてない時の解析と同様に t をそれぞれの成分に分けて解析を行ったところ、$t_{\mathrm{lag}} = 42$ 日と 51 日付近の解では、$w = 20$ 日付近の解はなく、$t_{\mathrm{lag}} = 46$ 日付近の解にのみ $w = 20$ 日付近の解があることが分かった。この違いを図 3.34 と図 3.35 に示す。図 3.34 では確かに $w = 20$ 日付近に解が存在するが、図 3.35 では $w = 20$ 日付近に解が存在しないことが確認できる。t_{lag} に 45 日から 47 日に事前分布を置いた時の光度曲線を図 3.36 に示す。図 3.32 と比べて低周波成分の補正がある分振幅の変動が小さくなっているが有意な変化は見られない。

表 3.2 に低周波成分の補正がない時と補正がある時のそれぞれの結果を示す。以上のように、電波の光度曲線に見られる低周波成分の補正の有無に依らず推定にはほとんど影響がないことがわかった。

表 3.2: 低周波成分の補正があるときとない時の各パラメータの推定値。

	σ	τ（日）	t_{lag}（日）	w（日）	A
低周波成分の補正なし	$0.3^{+0.2}_{-0.1}$	7^{+13}_{-3}	46^{+6}_{-5}	$< 2.2,\ 21^{+9}_{-7}$	$0.6^{+0.6}_{-0.2}$
低周波成分の補正あり	$0.3^{+0.1}_{-0.1}$	7^{+7}_{-2}	47^{+6}_{-6}	$< 5.4,\ 20^{+20}_{-10}$	$0.5^{+0.2}_{-0.3}$

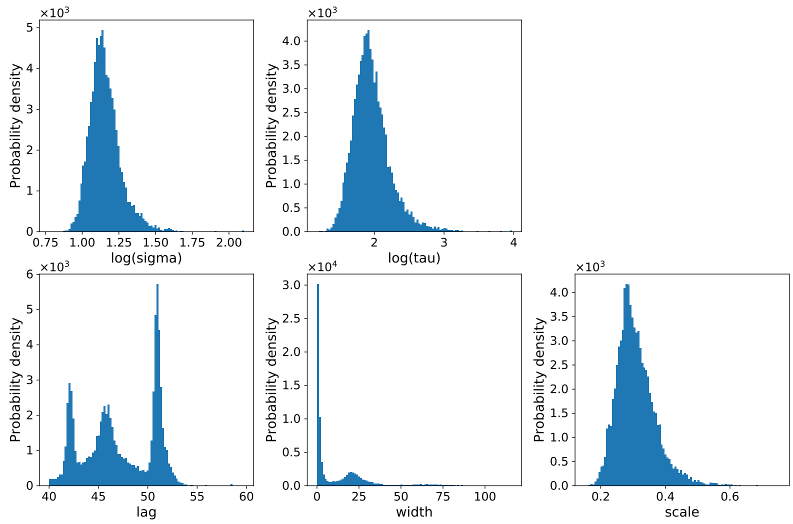


図 3.33: 低周波成分を引いた後の活動銀河核 PKS 1510−089 の t_{lag} に 40 日から 60 日の事前分布を置いた時の事後確率。横軸は上段左から σ、τ、下段左から t_{lag}、w、A である。縦軸は確率密度である。

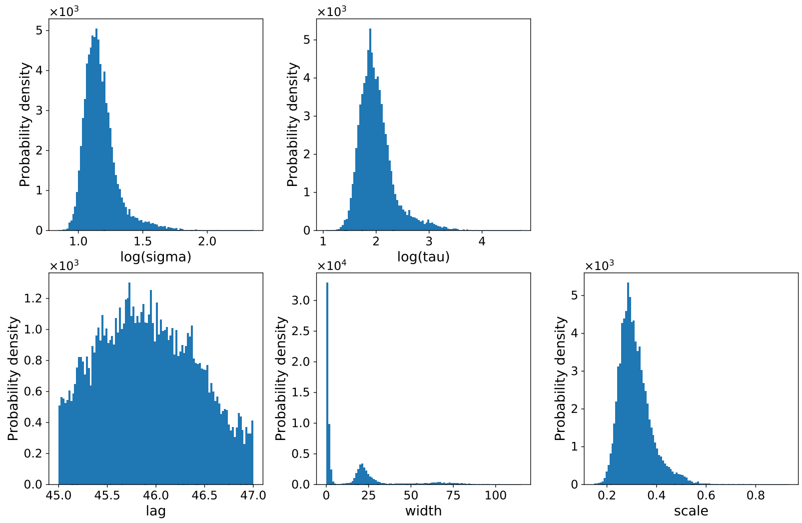


図 3.34: 低周波成分を引いた後の活動銀河核 PKS 1510−089 の $t_{\rm lag}$ に 45 日から 47 日の事前分布を置いた時の事後確率。横軸は上段左から σ、τ、下段左から $t_{\rm lag}$、w、A である。縦軸は確率密度である。

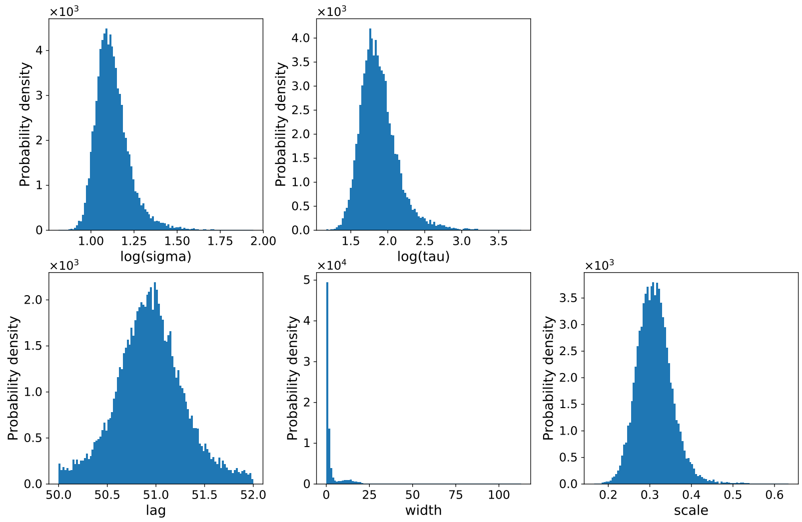


図 3.35: 低周波成分を引いた後の活動銀河核 PKS 1510−089 の $t_{\rm lag}$ に 50 日から 52 日の事前分布を置いた時の事後確率。横軸は上段左から σ、τ、下段左から $t_{\rm lag}$、w、A である。縦軸は確率密度である。

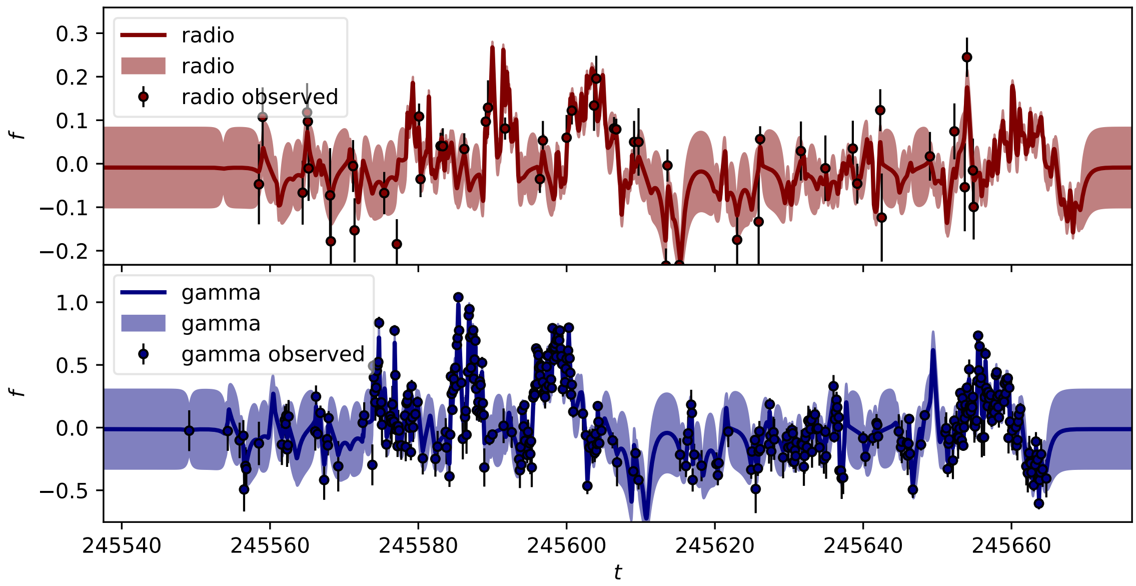


図 3.36: 低周波成分を引いた後の活動銀河核 PKS 1510−089 の t_{lag} に 45 日から 47 日の事前分布を置いた時の光度曲線。縦軸が光度、横軸が JD である。上が低周波成分を引いたあとの電波、下がガンマ線である。

第4章　考察

4.1　$w \geq 4$ で真値以外に局所解が見られる原因

3.1.2 章で t_{lag} の値に関わらず、w は 4 以上になると真値のほかに有意な局所解が見られることがわかった。この結果は w の推定の成否に t_{lag} は関係ないことを示している。w と相関があるパラメータとして考えられるのは、OU 過程の時間スケールである τ、人工データのパラメータの観測基線長、データ間隔である。σ や A は光度変動の振幅に関するパラメータなので相関があるとは考えにくい。また、誤差も時間変動とは全く関係ない人工データのパラメータなので、パラメータと相関があるとは考えにくい。そこでこの章では前述した 3 つのパラメータと w の関係を調べることで、$w \geq 4$ で真値以外の局所解が見られる原因を考察する。

始めに、w の成否に関係するパラメータとして考えられるのは τ である。τ は時間変動のタイムスケールを決めているパラメータであり、τ に対して w が大きいと畳み込み関数によって時間変動の情報が失われてしまうことが考えられるからである。本実験では、τ は 2、3、4、5、10、w は 1、2、3、4、5、10 と変化させ、w の推定の成否と τ の関係を調べた。その他のパラメータは既定値で行う。実験結果を表 4.1 に示す。表中の色の意味は表 3.1 と同様である。図中の青、緑の部分の事後確率の例は図 3.4 と同様であった。表 4.1 より τ を変化させても、3.1.2 の結果と同様に、$w \geq 4$ で一様に推定に失敗することが分かる。よって、w の推定の成否には τ の値は関係ないことが分かる。

次に w の推定の成否に関係するパラメータとして考えられるのは、観測基線長である。観測基線長をこれまでの実験では 100、それに合わせデータ数も 100 点で固定していたが、本実験では観測基線長を 200、それに合わせてデータ数も 200 点、w は 1、2、3、4、5、10 で実験を行う。その他のパラメータは既定値で行う。観測基線長を 200 とした時の、推定結果を表 4.2 に示す。表中には観測基線長が 100 の時の結果として表 3.1 の $t_{\mathrm{lag}} = 20$ の時の結果も示している。表中の色の意味は表 3.1 と同様である。図中の青の部分の事後確率の例は図 3.4、緑の部分の事後確率の例は図 3.5 と同様であった。表 4.2 から分かるように観測基線長を 200 にしても観測基線長が 100 の時と結果は変わらなかった。よって、w と観測基線長は関係がないことが分かった。

残りの w に関係するパラメータはデータ間隔だけである。他のパラメータが関与している可能性は低いと考えられるのでこのパラメータが w の推定に関係している可能性が高い。本論文を通じてデータ間隔は 1 に固定しているが、例えばデータ間隔を 0.5 に変えれば時間変動に関するパラメータは全て半分になることは自明である。w も 2 以上では真値以外に局所解がでることが容易に分かる。

実際、3.2 章で述べたように、PKS 1510−089 のデータ解析において、実時間スケールでは $w = 20$ 日の解は得られなかったが、時間スケールを 0.1 倍にすると、該当する解が得られた。これは、データ間隔が w の推定に影響を与えることを示唆する。このような挙動の原因は MCMC のアルゴリズムに関係する可能性もある。データ間隔の 4 倍以上の w の推定が難しい原因は今後も調査が必要である。

表 4.1: 様々な τ と w での推定の成否。横軸に τ、縦軸に w を取り、表中の色の意味は表 3.1 と同様である。

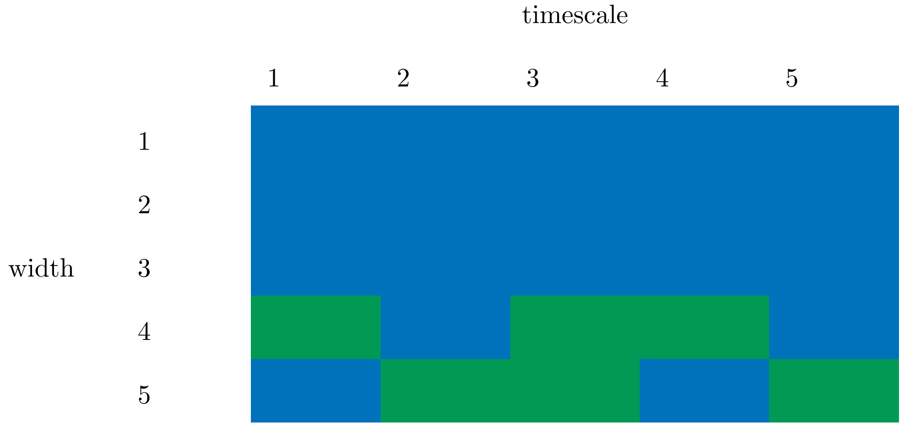
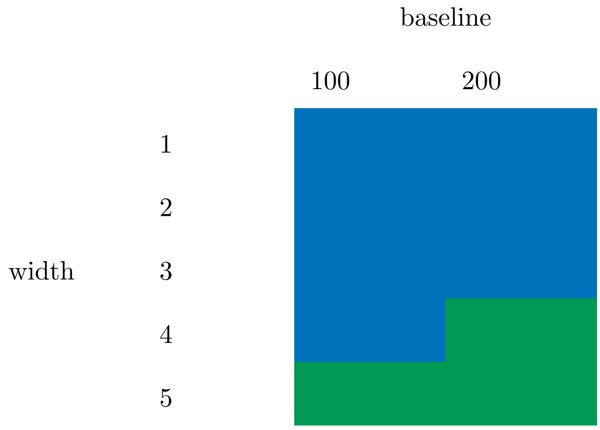


表 4.2: 観測基線長が 100 と 200 の w 推定の成否。横軸に τ、縦軸に w を取り、表中の色の意味は表 3.1 と同様である。



4.2　データの誤差と推定の成否

3.1.5 章で述べたように誤差が 1.0 までは w の推定に成功することが分かった。それよりも大きくなると推定に失敗することが多くなる。誤差が 1.1 の時の事後確率を図 4.1 に、光度曲線を図 4.2 に示す。誤差が 1.1 の場合は t_{lag} の推定には成功しているが、w の推定には失敗している例である。

図 4.1 を見ると w のピークが 1 付近と 5 付近にあることが確認できる。これは大きな測定誤差によって偶然明るくなった点を鋭いピークがあると捉えることもできるし、逆に偶然暗くなった点があることでフレアのピークが鈍っているとも解釈できる。例えばサンプル 2 の時間が 78 の時のデータに注目すると観測点は周辺において極値であるが、誤差も考えると周辺の観測点と同じ光度の可能性も否定できなくなる。もしこの光度が小さい場合は w の影響で鈍っていると解釈でき、光度が大きい場合は w の影響があまりない

鋭いピークであると解釈できる。つまり誤差が大きくなるとピークが鈍ることがあり、w と同じような働きをしてしまうので w の推定に失敗しやすくなることが考えられる。

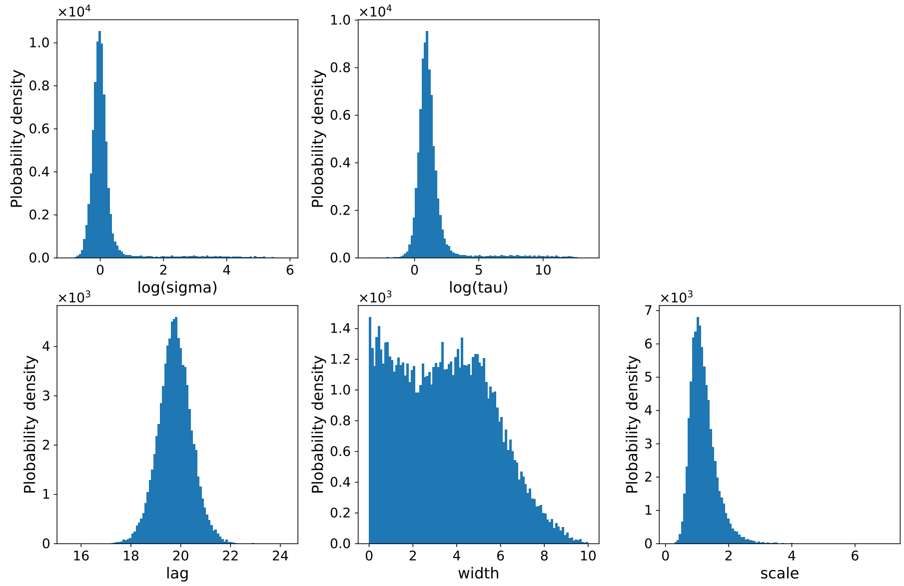


図 4.1: 誤差が 1.1 の時の事後確率。横軸は上段左から σ、τ、下段左から t_{lag}、w、A である。縦軸は確率密度である。

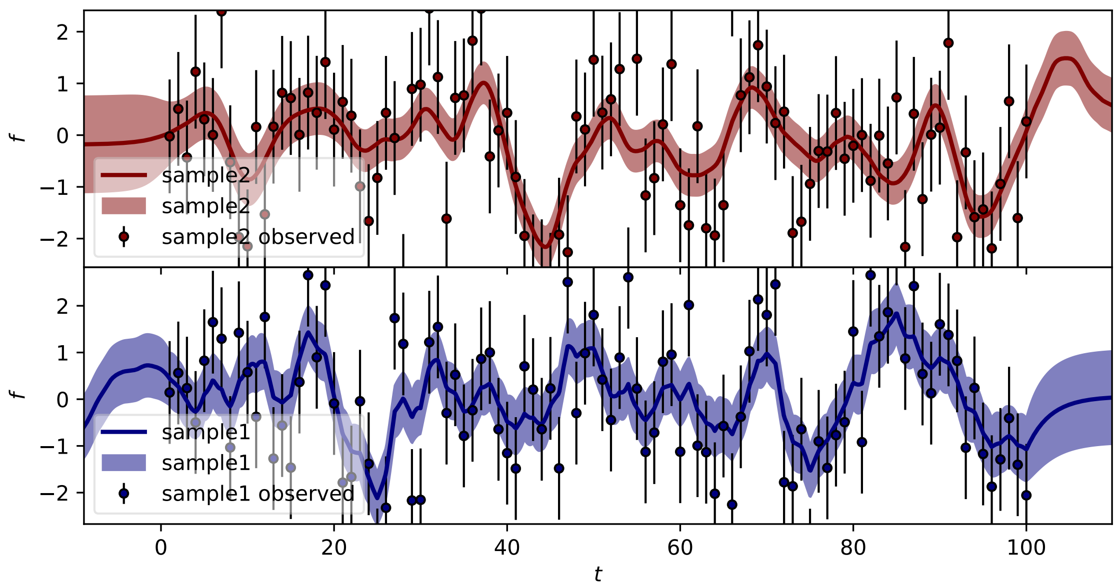


図 4.2: 誤差が 1.1 の時の光度曲線。

4.3 測定誤差が大きい場合の観測間隔よりも小さい t_{lag} の推定

この章では、観測間隔よりも小さい t_{lag} で推定に成功することについて考察していく。3.1.3 章で述べたように、観測間隔の 10%の t_{lag} の推定に成功することが分かった。これはサンプルが OU 過程に従って生成され、完全に OU 過程に従うとして推定しているからであると考える。観測間隔よりも小さい t_{lag} の推定は相関関数や DCF を用いた t_{lag} の推定では、相関係数を求める式の構造上推定することが不可能である。このことは従来の方法では見つけられなかった観測間隔よりも小さい t_{lag} の推定が、天体の時系列データが OU 過程に従うのなら可能だということを示唆している。一方で、3.1.3 章ではデータの誤差が無視できるほど小さい理想的な場合を考えていたため、実際のデータに応用する際に同じ結果が得られるかは自明ではない。そこで、この章ではより実際の天体の情報に近づけるために誤差が無視できない時の推定限界を調べることにする。

t_{lag} は、3.1.3 章で最小の t_{lag} である 0.1 で行う。誤差は 0.1 から 0.1 ごとに大きくして、t_{lag} が推定できる限界を調べる。w は 0、その他のパラメータは既定値で行う。この時の推定結果の成否を図 4.3 に示す。図 4.3 より、0.1 より小さい t_{lag} の推定は、誤差が 0.5 まで推定に成功し、それ以降は推定に失敗しやすくなることが分かる。

3.1.5 章では観測間隔よりも大きい t_{lag} で推定を行ない、誤差が 1.5 まで t_{lag} の推定に成功することが分かったが、本実験では誤差が 0.5 までしか t_{lag} の推定に成功しない。つまり、観測間隔よりも小さい t_{lag} の推定になると、誤差の影響を受けやすいことがわかる。

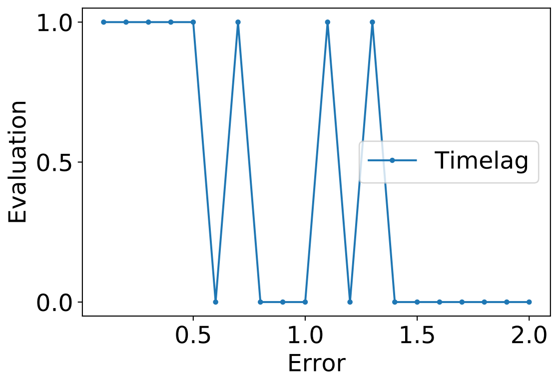


図 4.3: t_{lag} が 0.1 のサンプルに誤差を与えた時の t_{lag} の推定の成否。縦軸は 0 が推定に失敗、1 が推定に成功。横軸はデータに与えた誤差である。

4.4 欠損が多いデータを用いた推定

欠損データがある時の t_{lag} と w の推定の成否を 3.1.4 章で示したが全体の傾向として欠損がない完全なデータに対して少なくとも 3 割ほどデータが得られていると t_{lag}、w 両方の推定に成功することが分かった。

なぜ元々のデータ数の 3 割ほどで推定が可能なのか。それは、まず人工データを作る段階において OU 過程に完全に従うデータを作っており、次に推定において得られたデータが完全に OU 過程に従うとして

推定を行っているのでこれほどのデータ数で十分なのだと考えることができる。また、3.1.1 章から、2 割ほどデータが得られているならば、σ と τ の推定は可能だということが分かっている。このことも合わせると w の推定には、2 割では足りず 3 割必要であると考えることができる。一方で、欠損の多いデータでも推定に成功するのは、データの誤差が無視できるほど小さいためで、現実的な誤差が存在する場合の推定精度は低くなるかもしれない。また、3.1.4 章では一見してデータに相関が見られないほど少ないデータであっても推定に成功することを示したが、このことは逆に、全く相関関係がないデータでも偽の解が得られる可能性が示唆される。

そこで本章では、まず、4.3 章と同様に実際の天体の情報に近づけるために、欠損データに誤差がある時の推定の成否を実験する。次に、相関がないデータ間に偽の相関を与える可能性についても議論する。

4.4.1 欠損データとデータ誤差

この章では、欠損データに対して誤差を与えた時の推定の成否を実験する。一方のサンプルにランダムに欠損がありその数は 90 点、つまりデータ数は 10 点で、もう一方のサンプルは完全にある時で実験を行う。誤差は 0.1 から 0.1 ごとに大きくして、t_{lag} と w が推定できる限界を調べる。その他のパラメータは既定値で行う。この時の推定の成否を図 4.4 に示す。

図 4.4 より、t_{lag} は、誤差が 0.4 で推定に失敗しているものの最大で 0.6 までなら推定に成功することが分かる。また、w は誤差が最高で 0.3 までなら推定に成功することが分かる。

図 3.24 から分かるように、欠損がない完全なデータの場合は t の推定は誤差 1.6 まで、w の推定は誤差 0.9 まで成功する。これは図 4.4 と比べて明らかに大きい誤差で推定に成功している。したがって、欠損データが多い場合で、データの誤差が大きいと推定の精度が悪くなることが確かめられた。図 4.5 にサンプル 2 のデータ数が 10、誤差が 0.2 の時の事後確率を示す。図 4.5 を見ると、図 4.1 と同じように w の局所解が 2 つあることが確認できる。よって、欠損データが多く、誤差も大きくなると 4.2 章で考察したようにピークが鈍ってしまい、誤差が w と同じような効果を果たしていると考えることができる。

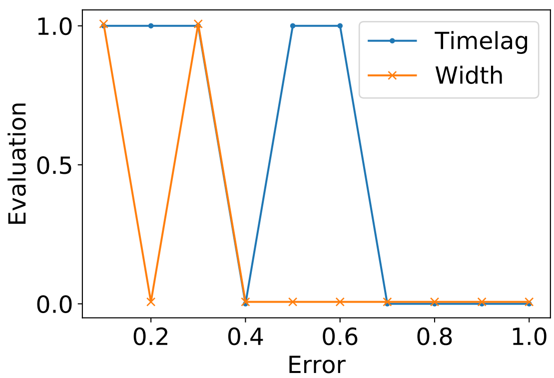


図 4.4: 10 点データに誤差を与えた時の t_{lag} と w の推定の成否。縦軸は、0 が推定に失敗、1 が推定に成功。横軸はデータに与えた誤差である。

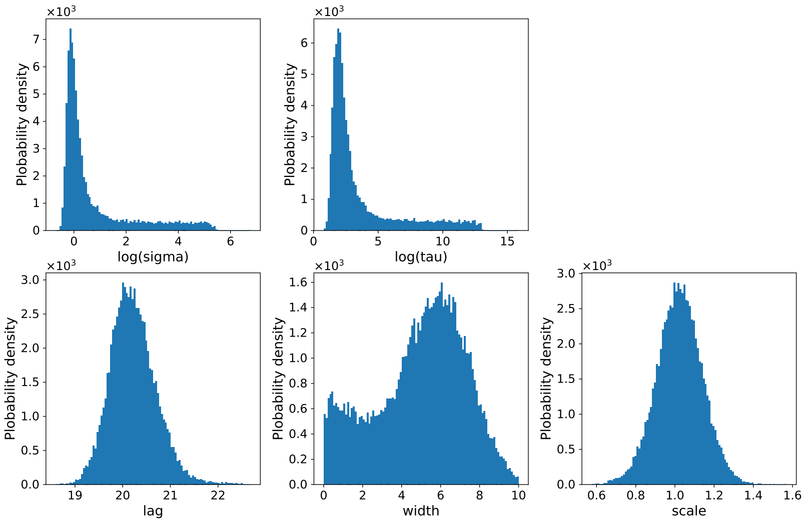


図 4.5: t の推定に成功、w の推定に失敗している時の事後確率。横軸は上段左から σ、τ、下段左から t_{lag}、w、A である。縦軸は確率密度である。

4.4.2 無相関なデータ間に相関を誤推定する可能性

3.1.4 章において、サンプル 1 が完全にあるならば、t_{lag} の推定にはサンプル 2 のデータが 5 点ほどで推定に成功することが確認できた。これほど少なくても推定できる理由として以下のことが考えられる。サンプル 1 のデータが σ や τ を推定するだけに十分な数があり、サンプル 1 のデータでこれら 2 つのパラメータの推定ができる。σ、τ は光度曲線のパラメータなので、これらを定めることで光度曲線が決まり、あとは応答関数のパラメータである t_{lag}、w、A を定めるだけになる。サンプル 2 はサンプル 1 と応答関数でモデル化されてるので、得られたサンプル 2 のデータに最も合うような応答関数を定めればよく、本研究の結果、それら 3 つの推定はデータが 5 点あれば可能であることが示された。

一方で、5 点というデータでは、実際の観測データの場合、光度変動が OU 過程にしたがっていると言うのは難しい。仮に、OU 過程に従う光度曲線データ 100 点と、それとは本来無相関な光度曲線データが 100 点ある時、それらの光度曲線間に有意なタイムラグと相関が見られる確率は極めて低いだろう。しかし、光度曲線に欠損が多く、5 点しかデータがない場合、それが偶然、相関するタイムラグの解が存在する確率は無視できないかもしれない。

そこでこの章では比較するサンプルに全く相関がないときに誤って相関を推定してしまう可能性について実験する。本実験では、サンプル 1 は OU 過程から生成したサンプル、サンプル 2 は正規乱数から生成したサンプルを用いる。サンプル 1 は 100 点、サンプル 2 は 10 点で行う。サンプル 1 のパラメータは既定値である。サンプル 2 の正規乱数の標準偏差は 1 で行う。

これまでの実験から w よりも t_{lag} の方が推定しやすいパラメータであることが分かっているので、本実験では、t_{lag} が -100 から 100 までの事後確率を求め、-50 から 50 の間に有意な局所解が存在していた時に、間違った相関があると判断する。試行回数は 100 回行う。

推定の結果、100 回のうち 20 回有意な局所解が見られた。あたかも t_{lag} があるように見える事後確率を図 4.6 に、t_{lag} の事前分布を 10 から 20 においた時の事後確率を図 4.7 に、光度曲線を図 4.8 に示す。図 4.6 と図 4.7 から、確かに t_{lag} が 15 付近に局所解があることが確認でき、図 4.8 では、ランダムに与えたサンプルにあたかも t_{lag} が生じているように光度曲線が回帰されていることが確認できる。

以上の実験より、データ数が少ない時には、実際には相関がないのにあたかも相関があるように推定してしまう可能性があることがわかった。

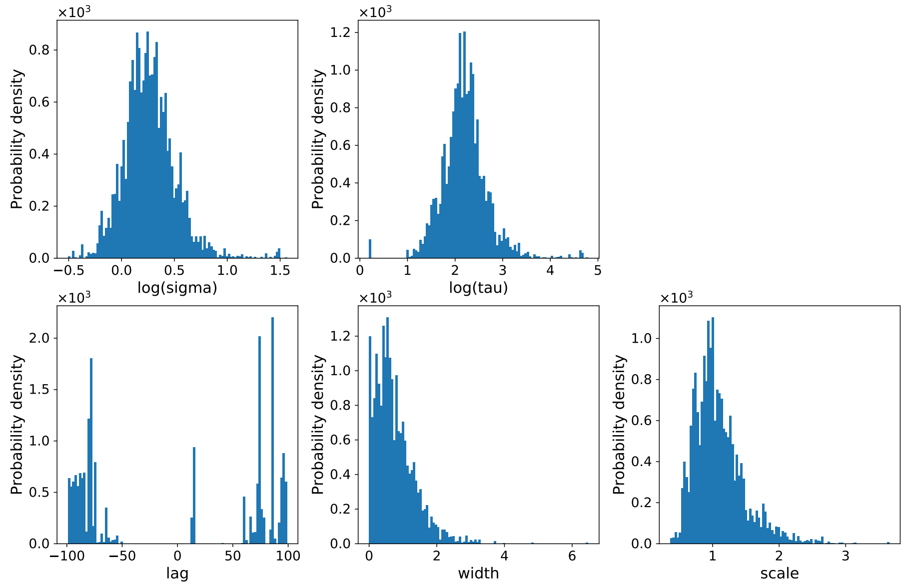


図 4.6: 相関がないサンプルを推定した時の事後確率。横軸は上段左から σ、τ、下段左から t_{lag}、w、A である。縦軸は確率密度である。

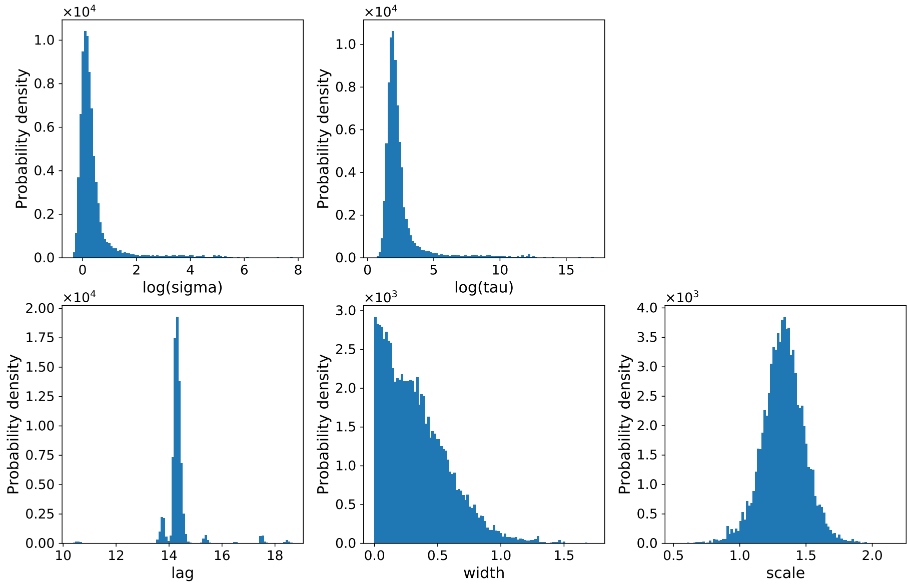


図 4.7: 相関がないサンプルを推定した時に t_{lag} に 10 から 20 の事前分布をおいた時の事後確率。横軸は上段左から σ、τ、下段左から t_{lag}、w、A である。縦軸は確率密度である。

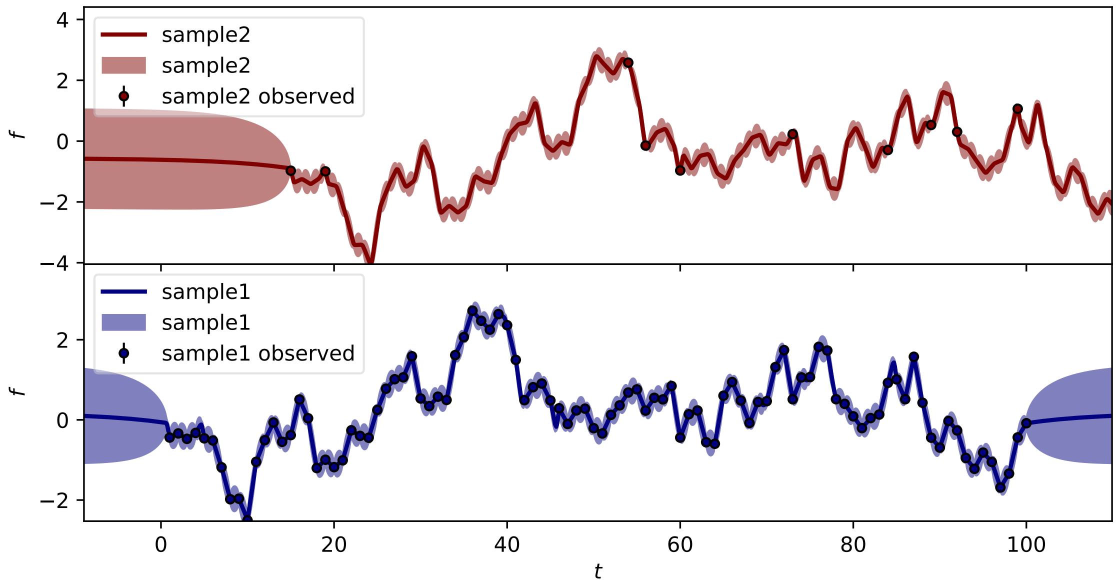


図 4.8: サンプル 2 にランダム欠損がありデータが 15 点の時の光度曲線。

4.5 活動銀河核 PKS1510−089 の推定結果から考えられる物理現象

活動銀河核 PKS 1510−089 のガンマ線と電波の光度曲線を解析した 3.2 章では、w の大局解は 0 日にあるが、20 日付近にも局所解があることが分かった。タイムラグに加えて、このタイムストレッチに相当する量は従来の相関係数を用いた手法では得られることのできなかった量で、ジェットの研究に新しい情報を加えて議論することができる。この章では、得られた結果をもとに考えられるジェットの構造を議論する。

PKS 1510−089 のようなブレーザー天体の場合、観測から推定される変動タイムスケールは以下の 2 つのタイムスケールが考えられる。

1. 放射領域の大きさに対して情報が伝わる時間スケール（light crossing timescale）
2. プラズマが放射冷却する時間スケール（cooling timescale）

ここでは 1 の light crossing timescale の描像を考える。この時考えられるイメージを図 4.9 に示す。中心天体に降着するガスが重力エネルギーを解放し、そのエネルギーの一部がジェットとして外に放出される。ジェットの中のプラズマは互いに衝突し衝撃波を形成するなどして、高エネルギーに加速された電子から電磁波が放射される。中心天体から放出されたプラズマはジェットの上流で電子が加速され、ガンマ線を放射する。図 4.9 に示すように、その時の放射領域の大きさを見積もるのに観測された変動から推定されるタイムスケールを用いることができる。放射領域がある大きさを持っているとき、その端から端まで情報が伝達する時間を OU 過程のタイムスケール τ とみなすことができる。このとき、ガンマ線の放射領域の大きさ R_γ は、ドップラー因子や赤方偏移を考慮すると式 (4.1) で表すことができる。

$$R_\gamma = \frac{Dc\tau}{1+z} \tag{4.1}$$

ここで D はビーミング因子、c は光速、z は赤方偏移である。

ガンマ線の上流でガンマ線を放射したプラズマは、ジェットの下流まで移動して電波を放射する。ガンマ線と電波の放射領域の距離は t_{lag} から見積もることができ、放射領域間の距離 $l_{\gamma-\mathrm{radio}}$ は次の式で表される (Janiak et al., 2012)。

$$l_{\gamma-\mathrm{radio}} = \frac{ct_{\mathrm{lag}}D\Gamma}{1+z} \tag{4.2}$$

ここで Γ はジェットのローレンツ因子である。電波の放射領域の大きさ R_{radio} は推定した w を用いて

$$R_{\mathrm{radio}} = \frac{Dc(\tau+w)}{1+z} \tag{4.3}$$

と考えることができる。この時、畳み込み関数によって時間変動が鈍る時の様子が図 4.9 に描かれている。ガンマ線の放射領域では τ の間ガンマ線を放射し、電波の放射領域では $\tau+w$ の間電波を放射する。

低周波成分の補正をしていない時の解析結果では、活動銀河核 PKS 1510−089 のガンマ線と電波の、$\tau = 7^{+13}_{-3}$ 日に分布の中心、$t_{\mathrm{lag}} = 46^{+6}_{-5}$ 日が確率が最も大きく、w は、0 日が 50%、20 日付近が 30%の確率であることが分かった。以上の結果を用いて放射領域間の距離や、放射領域の大きさを見積もることができる。典型的な値である、$D = 20$、$\Gamma = 15$、PKS 1510−089 の赤方偏移 $z = 0.36$ を用いて値を計算すると、放射領域間の距離は式 (4.2) より $l_{\gamma-\mathrm{radio}} \simeq 2.6^{+0.3}_{-0.3} \times 10^{19}$ cm、ガンマ線放射領域の大きさは式 (4.1) より $R_\gamma \simeq 3^{+5}_{-3} \times 10^{17}$ cm と考えることができる。電波の放射領域の大きさは $w \leq 2.2$ 日であればガンマ線と全く同じであり、$w = 21^{+7}_{-7}$ 日であれば式 (4.3) より $R_{\mathrm{radio}} \simeq 11^{+6}_{-3} \times 10^{17}$ cm であると見積もることができる。

電波の放射領域の大きさは、$w \leq 2.2$ 日の解の場合はガンマ線の放射領域と全く同じであることを意味する。これは、ジェット上流の $R_\gamma \simeq 3^{+5}_{-3} \times 10^{17}$ cm の大きさを持った領域でガンマ線を放射したあと、ジェット下流まで $l_{\gamma-\mathrm{radio}} \simeq 2.6^{+0.3}_{-0.3} \times 10^{19}$ cm 移動して、放射領域のサイズを全く変えないまま電波を放射したことになる。一方で、$w = 21^{+9}_{-7}$ 日の解の場合は、ジェット下流で $(\tau + w)/\tau = (20 + 7)/7 \sim 4$ 倍放射領域が大きくなったことになる。

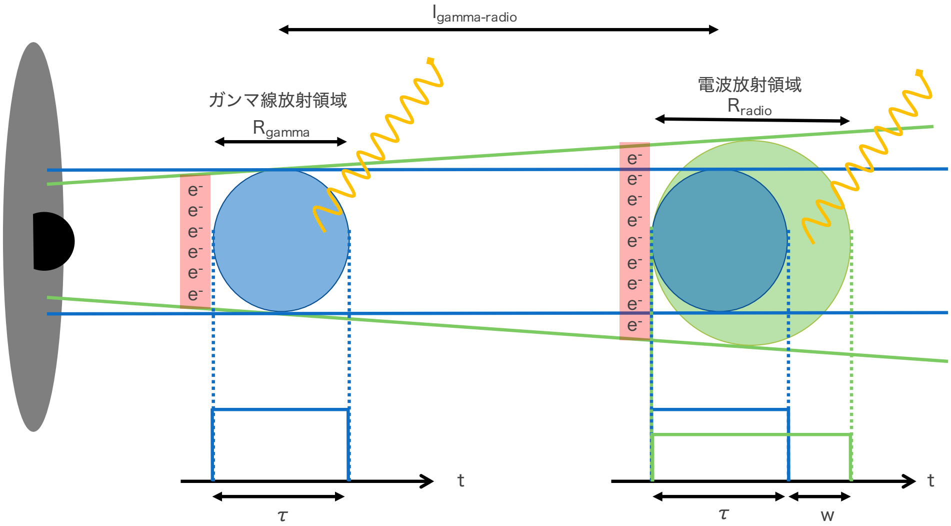


図 4.9: ジェットがガンマ線と電波の放射領域に達する時のイメージ図。

第5章　まとめ

本論文では確率過程モデルである OU 過程を用いて、ジェット天体の多波長同時観測された光度曲線のパラメータの推定方法を研究した。始めに人工データを用いて性能評価や推定限界を調べ、実際の天体の例として、先行研究でガンマ線と電波のタイムラグが 54 日と推定されている PKS 1510−089 のパラメータの推定を行った。人工データを用いた性能評価では以下のことが分かった。なおこれらはパラメータの既定値として、σ は 1、τ は 5、t_{lag} は 20、w は 2、畳み込み関数の振幅は 1、データ数は 100 点、誤差の標準偏差は 0.000005 とした時の結果である。

1. モデルのパラメータである OU 過程の分散パラメータと時間スケールは推定に失敗しやすいが、応答関数のパラメータのタイムラグ、タイムストレッチ、振幅の推定には影響がない。
2. タイムストレッチの推定には、タイムラグ、時間スケール、観測基線長は関係がないことがわかった。しかし、長いタイムストレッチは推定が困難である事がわかり、その原因は今後明らかにする必要がある。
3. 相互相関関数や DCF を用いた推定では見つけることができない、観測間隔よりも小さいタイムラグの推定に成功する。
4. データ欠損のパターン（ランダムに欠損もしくは一定間隔に欠損）に関わらず、3 割ほどデータが得られていればタイムラグとタイムストレッチの推定に成功する。
5. データの誤差が観測値の 100%程度の誤差であればタイムラグとタイムストレッチの推定に成功する。ただし、データに欠損がある場合で推定に成功するには誤差が小さい必要がある。

PKS1510−089 のデータを用いた推定では、先行研究と似た値のタイムラグの推定に成功し、タイムストレッチは 0 日に大局解があるものの 20 日付近にも局所解があることがわかった。これは低周波成分の補正の有無にも関わらず同様な結果が得られた。この結果からタイムストレッチが 20 日付近にあるとすれば、下流に向かうにつれてジェットが広がっている描像を考えることができ、ガンマ線と電波の放射領域間の距離は推定したタイムラグから $2.6^{+0.3}_{-0.3} \times 10^{19}$ cm、ガンマ線の放射領域の大きさは時間スケールから $3^{+5}_{-3} \times 10^{17}$ cm、電波の放射領域の大きさは時間スケールとタイムストレッチから $11^{+6}_{-3} \times 10^{17}$ cm と見積もることができた。

謝辞

研究を進めるにあたり、植村先生には大変お世話になりました。宇宙物理の基礎的なことや統計的手法など多くのことを学ばせていただきました。非常に感謝してます。

初めての研究室配属で分からないところがたくさんありましたが研究室の皆様が初めから丁寧に教えてくださったおかげで不自由なく学校生活を送る事ができました。ありがとうございます。特にゆりかさんには植村先生の姉弟子として、宇宙、統計、先生の好みなど本当に多くのことを教えてもらいました。ゆりかさんのご指導なしではここまでスムーズに研究を進める事ができなかったと思います。本当にありがとうございます。ただ、りゅうやとのしょうもない会話に巻き込んでしまいご迷惑をおかけしました。この場をお借りして謝りたいと思います。ごめんなさい。

2019 年 2 月 B209 にて　大間々知輝

2017 年 8 月 19 日宮崎県日向市金ヶ浜にて

参考文献

Beaklini, P. P., Dominici, T. P., & Abraham, Z. 2017, Astronomy & Astrophysics, 606, A87

Bishop, C. M. 2012, パターン認識と機械学習　下　ベイズ理論による統計的学習 (丸善出版社)

Edelson, R. & Krolik, J. 1988, The Astrophysical Journal, 333, 646

Gandhi, P., Bachetti, M., Dhillon, V. S., Fender, R. P., Hardy, L. K., Harrison, F. A., Littlefair, S. P., Malzac, J., et al. 2017, Nature Astronomy, 1, 859

Janiak, M., Sikora, M., Nalewajko, K., Moderski, R., & Madejski, G. M. 2012, The Astrophysical Journal, 760, 129

Koyama, K. & Minesige, S. 2007, ブラックホールと高エネルギー現象. Vol. 8 of シリーズ現代の天文学 (日本評論社)

Kozłowski, S. 2017, Astronomy & Astrophysics, 597, A128

Mirabel, I. & Rodriguez, L. 1998, Nature, 392, 673

Peterson, B. M. 1993, Publications of the Astronomical Society of the Pacific, 105, 247

Shibata, K., Fukue, J., Matsumoto, R., & Minesige, S. 1999, 活動する宇宙 -天体活動現象の物理- (裳華房)

Urry, C. M. & Padovani, P. 1995, Publications of the Astronomical Society of the Pacific, 107, 803

Zu, Y., Kochanek, C., & Peterson, B. M. 2011, The Astrophysical Journal, 735, 80