

粗視化スケールパラメータを用いた時系列解析 : ノイズ除去効果の定量的検証

大塚 史子 ・ 羽田 亨 (九大・総理工)

1. はじめに

レビ分布は、スケール不変な分布として重要である。この分布は大きなスケールでべき指数 3 以下の裾野の広いべき乗則に従う。そのため、この分布に従う粒子の拡散過程は非古典的となることが知られている。ここで非古典的拡散とは、粒子の分散が経過時間のべき乗に比例し、古典的拡散係数 $D = \langle \Delta r(\tau)^2 \rangle / \tau \sim \tau^\beta$ ($\beta \neq 0$) と記述される場合である。

われわれはこれまで、時間的に定常な磁場ゆらぎ中での荷電粒子の拡散過程をテスト粒子計算により議論してきた¹⁾。Fig. 1 に、2 次元磁場ゆらぎ中での粒子拡散の数値計算の結果を示す。図 (a) は、ある時間での粒子の移動距離の二乗平均 $\langle \Delta r^2 \rangle$ より、古典的拡散係数 D を経過時間の関数として評価している。図 (b) は、典型的な粒子の案内中心の時系列である。今、粒子のラーマ半径と磁場ゆらぎの相関長が同程度であるので、粒子は磁場等高線を乗り換えることができる。そのため、粒子は磁場等高線が形成する磁場アイランドによる捕捉 (停滞)、非捕捉 (歩行) を繰り返し、古典的拡散の代表であるブラウン運動とは異なる時系列となっていることがわかる。そして、このような粒子のアンサンブル平均より得られる拡散係数 D は、 $10^2 < t < 10^4$ 程度の時間スケールにおいて $\beta < 0$ の準拡散となっていることがわかる。

非古典的な拡散過程はさまざまな分野で観測されている²⁾。Weeks らは、渦とジェット流が形成された回転水槽中で、レビ歩行するトレーサー粒子の運動を観測した。トレーサー粒子は、渦による捕捉や、ジェット流による輸送を繰り返し、レビ歩行を行う。彼らは、これらの粒子軌道の時系列を停滞と

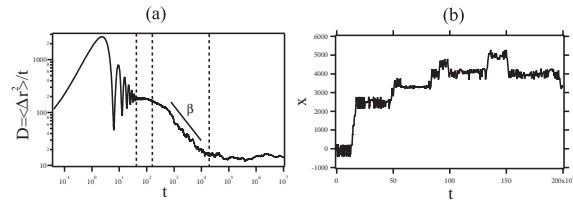


Fig. 1 Test particle simulation result of cross-field diffusion of charged particles : (a) diffusion coefficient vs. time, (b) particle trajectory

歩行の時間間隔から成ると考え、停滞と歩行時間間隔の確率分布関数 (p.d.f.) をべき乗と仮定してモデルを作った。停滞と歩行時間間隔の p.d.f. のべき指数の大小関係により、拡散過程は超拡散 ($\beta > 0$)、準拡散 ($\beta < 0$) となる。彼らは、べき指数 β と p.d.f. のべき指数との間のスケーリング則を理論的に評価した³⁾。

レビ分布に特徴づけられる粒子軌道の時系列と拡散係数の時間スケールのべき指数 β を関連付けるためには、時系列を停滞と歩行の時間間隔に分離する必要がある。その際、Fig. 1(b) に見られるような、数値計算あるいは観測によるデータには、停滞と歩行の時間間隔からなる時系列以外のノイズの存在は必須であり、ノイズの除去が必要である。本稿では、ノイズ、データの時間離散幅の影響を考慮し、時系列を停滞と歩行時間に分離するアルゴリズムを考える。粗視化スケールパラメータ λ を用い、時系列分離に対し有用なパラメータを検討する。

2. レビウォーク

ここでレビ統計とはどのようなものか、簡単に紹介する。レビ分布とは以下のように記述される分布

である。

$$p_N(k) = e^{-N|k|^\gamma} \quad (0 < \gamma \leq 2) \quad (1)$$

$$p_N(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} dk p_N(k) e^{ikx} \quad (2)$$

このレビ分布にはいくつかの特徴がある。その重要なひとつに、スケーリング則が成り立つことがあげられる。すなわち、

$$p(x) = N^{1/\gamma} p_N(xN^{1/\gamma}) \quad (3)$$

と書き表すことができる。ここで、添え字なしの分布 $p(x)$ は、(2) 式において $N = 1$ とした場合である。添え字の N は $p(x)$ を N 回畳み込み積分することを意味する。その分布 $p_N(x)$ は、 $p(x)$ と (3) 式のようなスケール変換が成り立ち自己相似性を表す。

このようなスケール不変な分布はガウス分布が代表的であるが、より一般的なスケール不変の分布として P. Lévy(1937) により (1) 式のレビ分布が発見された。(2) 式において、 $\gamma = 2$ とすることにより、分散 = $2N$ のガウス分布が得られる。また、 $\gamma = 1$ で、コーシー分布となる。

さらに、 $x \rightarrow \infty$ で以下のようなべき乗則に従う。

$$p(x) \sim |x|^{-\gamma-1} \quad (4)$$

ここで、この分布に対し、2次のモーメント $\langle x^2 \rangle$ を考えると、 $\gamma < 2$ で発散してしまうことがわかる。しかし、仮に等速直線運動する粒子を考え、その粒子はある距離まで移動すると逆向きに移動する、というような運動(歩行)をランダムに繰り返すとする。ただし、その粒子の一回の移動距離 $x = vt$ が (4) 式に従うとする。そのときの経過時間 t の粒子の二乗平均は、

$$\langle x(t)^2 \rangle = \int_{x_0}^{vt} x^2 p(x) dx + \int_{vt}^{\infty} (vt)^2 p(x) dx \quad (5)$$

と考えることができる。ここで簡単のため最小スケール x_0 を導入した。右辺の第一項は、観測する時間 t に移動できる距離 $x = vt$ よりも短い移動距離を持つ粒子、第二項目は $x = vt$ より長い移動距離を持つ粒子に対応する。同様に、平均的な移動時間 $\langle t \rangle = \langle x(t) \rangle / v$ として計算し、拡散係数を $D = \langle x(t)^2 \rangle / \langle t \rangle$ と考える。時間が十分に大きい場合 $t \rightarrow \infty$ を考えると、

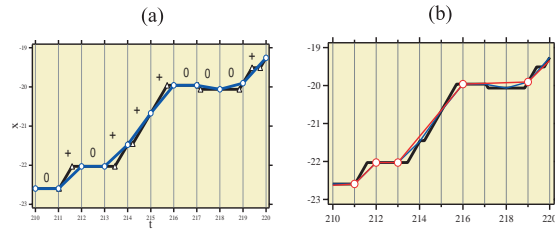


Fig. 2 (a) typical test data produced by levy distribution. (b) data obtained by separation algorithm using data(a)

$$D \sim const \quad (\gamma > 2) \quad (6)$$

$$D \sim t^{2-\gamma} \quad (1 < \gamma < 2) \quad (7)$$

$$D \sim t \quad (0 < \gamma < 1) \quad (8)$$

を得る。拡散過程は $\gamma > 2$ で古典拡散、 $1 < \gamma < 2$ のときに超拡散、 $\gamma < 1$ で等速直線運動であることがわかる。 $\gamma < 2$ では、2次のモーメントが発散することを反映し、有限の時間スケールで粒子の二乗平均を観測することにより、非古典的拡散過程があらわれる。この例では、粒子の移動距離がレビ分布に従うとしたが、これは、歩行の時間間隔がレビ分布に従うと考えることと同じである。さらに、停滞の時間間隔も存在し、その p.d.f. がレビ分布に従うとすると、準拡散があらわれる。前述した Weeksらのモデルがこの場合である。

3. 解析方法

時系列を停滞と歩行の時間間隔に分離するアルゴリズムを考えるにあたって、テストデータを生成する。まず、レビ分布に従う、停滞と歩行時間間隔のランダム変数⁴⁾を交互に選択する。移動距離は $\Delta x = vt$ であり、 $v = 0, \pm 1$ である。ここで歩行の速度の符号はランダムに選択する。この移動距離を足し合わせて時系列を作る。ここでのパラメータは、分布のべき指数 γ と標準偏差 σ である。なお、(2) 式の N を用い $\sigma = (2N)^{1/\gamma}$ である。停滞と歩行それぞれについて、べき指数は $\gamma_s = 2, \gamma_w = 1.5$ 、標準偏差は $\sigma_s = 1, \sigma_w = 1$ と固定する。

Fig. 2(a) の印のデータがこのようにして作られた典型的なデータである。これは理想的な停滞と歩行からなる時系列であるが、実際のデータは離散的であるため、図の印のデータになる。この例ではデータ離散幅 $\Delta t = 1$ である。さらに、この印

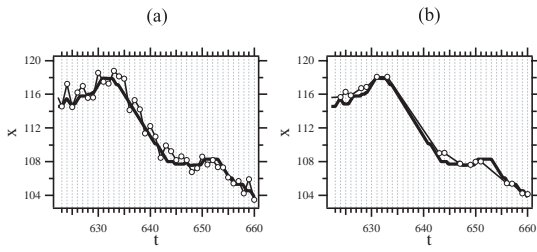


Fig. 3 data obtained by algorithm using data with noise. (a) $\lambda = 2$. (b) $\lambda = 4$.

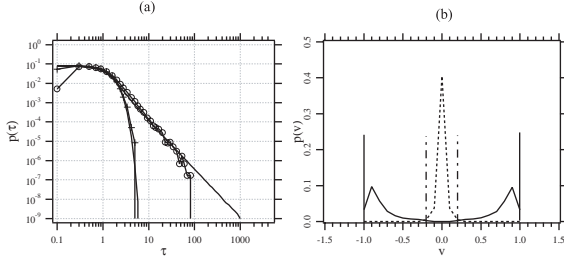


Fig. 4 p.d.f. without noise ($\lambda = 2, v_c = 0.2$). (a) p.d.f. of stick and walk duration time. (b) p.d.f. of averaged velocity.

のデータ各点に標準偏差 σ_{noise} のガウスノイズを加えたデータを作成する。

この離散的なデータを用い、時系列を再び停滞と歩行間隔に分離する。粗視化スケールパラメータ λ を用い、各点での傾きを以下のように定義する。

$$v_j^\lambda = \frac{x_{j+\lambda/2} - x_{j-\lambda/2}}{\lambda \Delta t} \quad (9)$$

定義された傾き v_j^λ に対し、速度閾値 v_c を用い

$$|v_j^\lambda| > v_c \rightarrow \text{Walk (+, -)} \quad (10)$$

$$|v_j^\lambda| < v_c \rightarrow \text{Stick (0)} \quad (11)$$

と判定する。判定した例を Fig. 2(a) に示す。同じ符合がつながっている間隔をひとつの停滞または歩行と考え、その時間間隔は $\tau = \Sigma \Delta t$ とする。ひとつの時間間隔に対し、速度の平均 $\bar{v} = \langle v_j^\lambda \rangle$ を求め、移動距離は $\Delta x = \bar{v} \tau$ となる。分離アルゴリズムのパラメータは離散幅 Δt 、粗視化スケールパラメータ λ および速度閾値 v_c である。このアルゴリズムを用いた結果を次に示す。

4. 計算結果

4.1 ノイズがない場合

ノイズがない ($\sigma_{noise} = 0$) 場合、停滞と歩行の時間間隔の確率密度関数 (p.d.f.) と、速度の確率密度関数を Fig. 4 に示す。 $\lambda = 2, v_c = 0.2$ である。(a) の p.d.f. の実線はそれぞれレビ分布 $\gamma = 1.5$ と $\gamma = 2$ (ガウス分布) である。 が歩行時間の p.d.f.、+ が停滞時間の p.d.f. である。これらは、与えたレビ分布とほぼ一致しているが、いくつかの違いがあげられる。まず、データの離散幅 $\Delta t = 0.1$ であるので、 Δt より短いスケールは分離することができない。また、今、データ長 $T = 100$ としたため、それ以上の分布は存在しないことがわかる。

次に速度の p.d.f. を見る。理想的な場合の歩行の速度の位置を $v = \pm 1$ に棒で示す。それ以外の実線が歩行の p.d.f.、破線が停滞の p.d.f. である。一点破線で示した棒は、速度閾値 $v_c = 0.2$ を示し、それを境に歩行と停滞が分けられている。理想的には、 $v = 0, \pm 1$ に分布が来るはずであるが、分布に幅があり、歩行の p.d.f. のピークは $v = \pm 1$ より内側にある。これは、Fig. 2(b) の分離された印のデータからわかるように、離散幅が有限なため、実際の速度よりも小さく見積もられることがわかる。 $\Delta t \rightarrow 0$ の極限では、分布の幅は狭く、ピークは $v = 0, \pm 1$ に漸近することが期待される。ここで、 $v_c = 0.2$ と選んだのは、ノイズがないため傾き $v > 0$ ならば、歩行であるため v_c を小さく選んだが、これも Δt に依存する。

ノイズがない場合に、分離が可能であった速度閾値 $v_c = 0.2$ を用い、ノイズがある ($\sigma_{noise} = 0.1$) 場合の解析を行った結果が Fig. 5 の (a)(b) である。(a) の $\lambda = 2$ では、時間間隔の p.d.f. は全くあっていない。また、速度の p.d.f. では、歩行の分布が広がっており、Fig. 3(a) で見られるようにノイズを拾うと傾きの分布に幅がでてくることがわかる。(b) の $\lambda = 8$ では、歩行の分布のテイルの部分が抽出されているが、停滞の分布の広がり小さい。速度の p.d.f. では、歩行のピークが停滞の分布に近く、あまり分離されていないことがわかる。これよりも λ を大きくすると、 λ 以下のスケールが合わなくなってくる。そこで、 $v_c = 0.4, \lambda = 8$ として解析した結果が (c) である。(c) は速度閾値 $v_c = 0.4$ とすることにより、(b) より時間間隔の分布は合う。速度の p.d.f. で見ると、 v_c を大きくした分歩行のピークは停滞の分布から離れているが、停滞の分布の端に

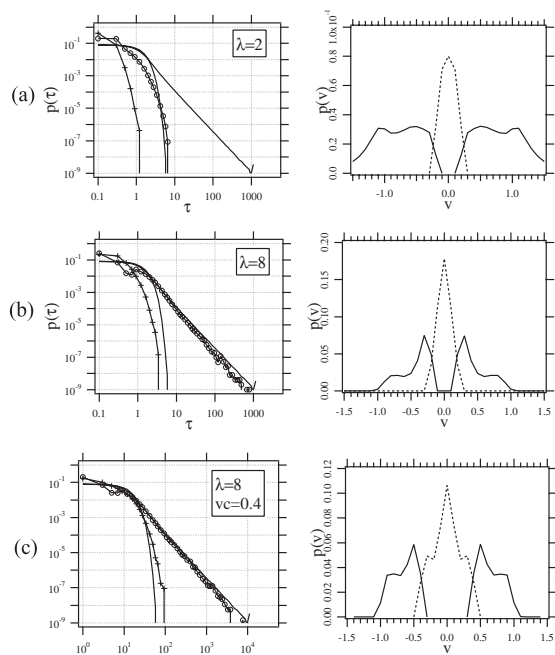


Fig. 5 p.d.f. with noise ($\sigma_{noise} = 0.1$) (a) $\lambda = 2, v_c = 0.2$. (b) $\lambda = 8, v_c = 0.2$.,(c) $\lambda = 8, v_c = 0.4$.

ピークが現れ、歩行の分布を拾っている傾向にある。

5. まとめ

MHD 乱流中の粒子運動は、停滞と歩行の組み合わせで記述できるような非ブラウン運動である。粒子アンサンブル統計の時間発展も、非古典的な枠組み（レビ統計）で考えることができる。与えられた粒子運動の時系列を停滞と歩行時間間隔に分離し、それぞれの確率密度関数 (p.d.f.) を求めるアルゴリズムを考えた。得られる p.d.f. の、元の p.d.f. からずれる要因と、それにより p.d.f. にどのような誤差が生じるかを以下にまとめる。

1. データ離散幅 $\Delta t \neq 0$
時間間隔の p.d.f. を Δt 以下は表現できない。
速度分布に幅が生じる。
2. データ長 $T \neq \infty$
時間間隔の p.d.f. が T までしか表現できない。
3. ノイズ $\sigma_{noise} \neq 0$
時間間隔の p.d.f. を再現できない。
速度分布に幅が生じる。

1 と 3 を解消する粗視化スケール λ と v_c の組み合わせは、今回は、速度分布をみて、ノイズがない場合の分布に近づくように、いろいろなパラメータ

に対し試すという方法をとった。このアルゴリズムをそのような補正が行えるように改良し、時系列データを最適に分離できる方法を今後考える。カイ二乗検定などを行い、p.d.f. の誤差を定量的に評価する。

このアルゴリズムを数値計算や、観測により得られたデータの解析に応用する。

参考文献

- 1) F. Otsuka, T. Hada, Cross Field Diffusion of Cosmic Rays in a Two-Dimensional Magnetic Field Turbulence, Space Science Reviews, in press (2003) .
- 2) J. Klafter, M.F. Shlesinger, G. Zumofen, Beyond Brownian Motion, Physics Today, Feb., 33 (1996).
- 3) E.R. Weeks, J.S. Urbach, H.L. Swinney, Anomalous diffusion in asymmetric random walks with a quasi-geostrophic flow example, Physica D, 97, 291 (1996) .
- 4) J.M. Chambers, C.L. Mallows, B.W. Stuck, A method for simulating stable random variables, Journal of the American Statistical Association, 71,340 (1976).