

3次元軌跡データの基本モデルとその限界

統計数理研究所 島谷健一郎

The basic model for three-dimensional animal trajectory data and its limitations

The Institute of Statistical Mathematics Ichiro Ken Shimatani

1. 序

動物装着型データロガーや高速カメラなどの技術の発達により、動物が動いた詳細な3次元軌跡データが得られる時代となっている。ここでは、その軌跡の特徴を抽出するのに便利な統計モデルを、2次元の軌跡データを自然に拡張する形で提唱する。

2次元から3次元への拡張は、 (x, y) を (x, y, z) に直すだけ、極座標で表すなら、 (r, θ) にpitch角を追加して (r, θ, φ) とするだけ、である。そこに確率論的不確実性を加える場合でも、 (x, y, z) 系ではすべてに正規分布誤差を加える形で済ませられる。

一方、極座標では、方向 θ について、円周上の確率分布が必要になる。3次元では、 (θ, φ) が従う球面上の確率分布が必要となる。方程式も、2次元から拡張しやすい形と、拡張しにくそうな形がある。ここでは、その最も拡張しやすい例を提唱し、実際の動物の3次元軌跡への適用例を示す。

2. 2次元の軌跡データモデル

2次元軌跡データについて、時刻 t での位置を \mathbf{X}_t 、 t から $t+1$ にかけて移動した方向を $\theta_t = \arg(\overrightarrow{\mathbf{X}_t \mathbf{X}_{t+1}})$ とする。2次元の場合によく使われる θ_t に関する自己回帰型モデルには次のようなものがある。

$$\theta_t = \alpha + 2 \tan^{-1}(w \tan(\theta_{t-1} - \alpha)/2) + e_t \quad \text{Kato (2010) 円周自己回帰モデル}$$

$$\theta_t = \arg(\cos(\theta_{t-1}) + w \cos(\alpha), \sin(\theta_{t-1}) + w \sin(\alpha)) + e_t \quad \text{von Mises process Breckling (1989)}$$

$$\theta_t = \alpha + \sin(\theta_{t-1} - \alpha)/w + e_t \quad \text{Kuramoto model } d\theta/dt = \sin \theta/w \text{ の離散化}$$

いずれも、次に進む方向は、1つ前の方向の維持を基本に、そこから目標方向 α へ軌道修正する形になっている。パラメータ w は方向 α へ向かおうとする強さを表す(動物

がその方向に行きたいという意志の強さとも解釈できる)。 e_t は確率論的誤差で、次節に示す円周上の確率分布からのランダムなサンプルである。

上に挙げた3つのモデルの中で、3次元に拡張しやすいのは、2番目の von Mises process である。

3. 球面上の von Mises 分布

その前に、円周上の von Mises 分布

$$f_{vM}(y; \mu, \kappa) = \exp(\kappa \cos(y - \mu)) / 2\pi I_0(\kappa)$$

は (μ は平均方向、 κ はその周りの集中度を示すパラメータ) 、

$$f_{vM}(y; \mu, \kappa) = \exp(\kappa(\cos y \cos \mu + \sin y \sin \mu)) / 2\pi I_0(\kappa) = \exp(\kappa(\vec{y} \cdot \vec{\mu})) / 2\pi I_0(\kappa)$$

と書き直すことで、次のように3次元に拡張できる (von Mises-Fisher distribution, または Fisher distribution、Mardia & Jupp 2000) 。

$$f_{vM}^3(\vec{y}; \vec{\mu}, \kappa) = \kappa \exp(\kappa(\vec{y} \cdot \vec{\mu})) / \sinh(\kappa)$$

これを確率密度関数とする球面上の確率分布を

$$vM^3(\vec{\mu}, \kappa)$$

で表すことにする。

4. 3次元の方向自己回帰モデル

von Mises process を、 $\overrightarrow{\mathbf{X}_{t-1}\mathbf{X}_t}$ を既に単位ベクトルに正規化されたベクトルとみなして、

$$\theta_t \sim vM(\arg(\cos(\theta_{t-1}) + w \cos(\alpha)), \sin(\theta_{t-1}) + w \sin(\alpha)), \kappa) = vM(\overrightarrow{x_{t-1}x_t} + w\vec{\alpha}, \kappa)$$

と書き直せば、確率分布を球面上の Fisher 分布に直すだけで3次元の方向自己回帰モデルとなる。

$$\theta_i \sim vM^3(\overrightarrow{x_{i-1}x_i} + w\vec{\alpha}, \kappa)$$

$\overrightarrow{X_{i-1}X_i}$ だけでなく $\overrightarrow{X_{i-2}X_{i-1}}$ にも依存させるなら

$$\theta_i \sim vM^3(\overrightarrow{x_{i-1}x_i} + w_{i-1}\overrightarrow{x_{i-2}x_{i-1}} + w_\alpha\vec{\alpha}, \kappa)$$

とすればよい。 w_α は3次元の(球面上の)方向 α へ向かう強さ(動物の意志の強さ)で、1つ前の方向を維持させる強さとの相対値である。 $w_{i,2}$ は2つ前の方向の影響で、 $w_{i,2} = -0.5$ は等角速度運動に対応する。 κ は1つあるいは2つ前に進んだ方向と目標方向のウェイト和で決まる次の方向へどのくらい正確に向かうかを示し、動物の方向調整能力、あるいは行きたい方向に向かおうとする意志を妨げる外部ノイズと解釈できる。

5. シミュレーション例

図1にいろいろなパラメータ値の元でシミュレーションで描いた例を示す。

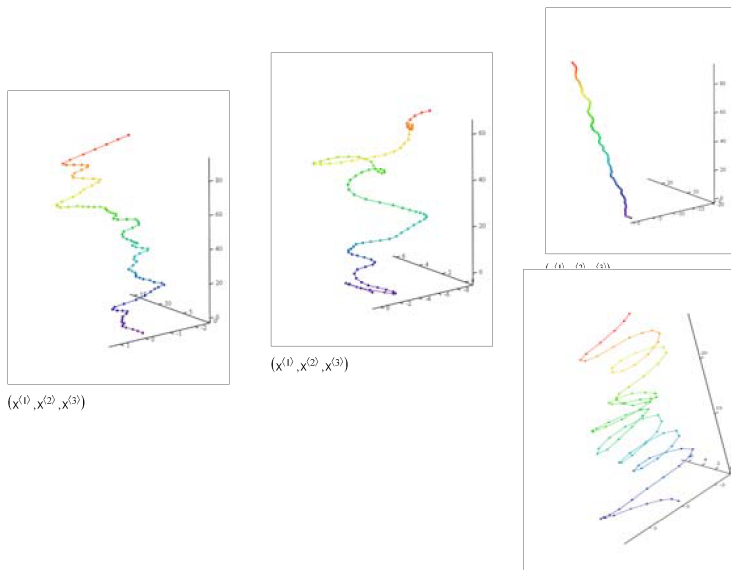


図1 3次元に拡張された von Mises process で作成した3次元軌跡の例

6. 実データへの適用例

ペンギンの潜水データを、深度変化から潜水、海底での徘徊、浮上、の3つに分け、それぞれに球面上の自己回帰モデルを適用し、パラメータを最尤推定し、最尤推定値を用いた軌跡例をシミュレーションで作成し、さらに残差を球面上にプロットし、球面上の確率分布からのランダムサンプリングと比較したところ、以下のようになった。

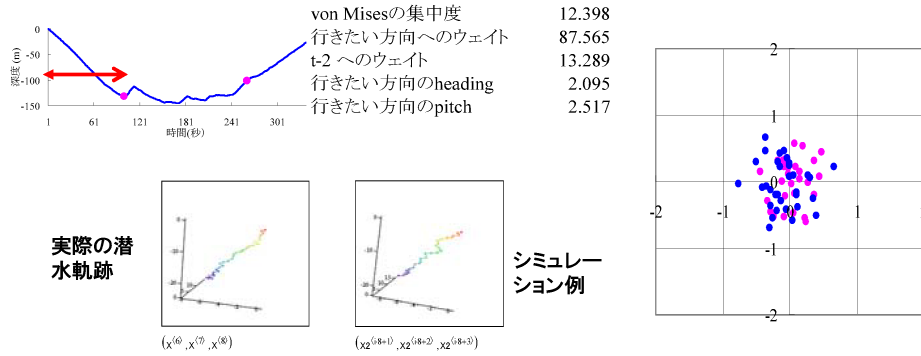
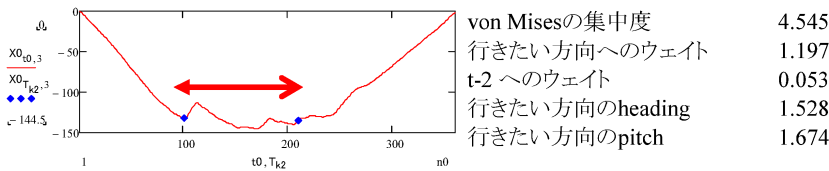


図2潜水。右の散布図は、球面の北極を原点、赤道を単位円、南極を半径2の円周、北半球を単位円板内、南半球を半径1と2の円周の間の円環領域、に対応させて表したもの。青は実データの残差、ピンクはFisher分布からのランダムサンプルを示す。



深度変化の実測とシミュレーション

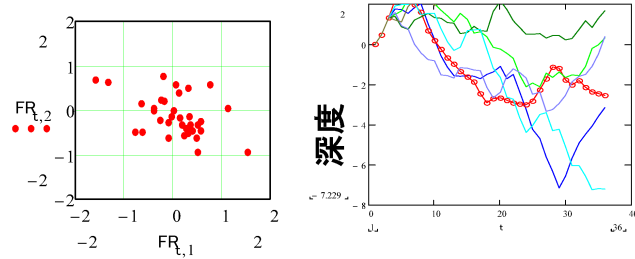
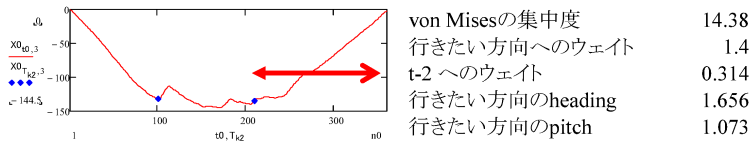


図3 海底での徘徊。左下の散布図は球面上の残差。



深度変化の実測とシミュレーション

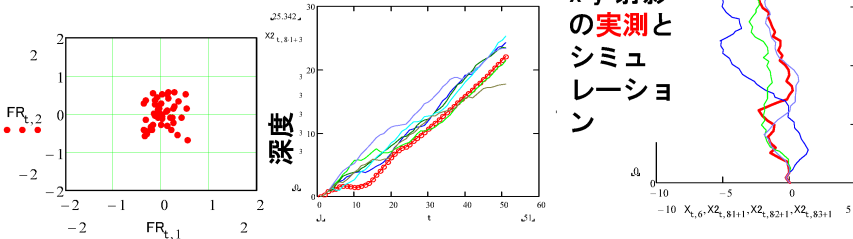


図4 浮上。左下の散布図は球面上の残差。

行きたい方向へのウェイトが潜水で高いことは、潜る行く先がほぼ決まっていることを意味するのかもしれない。Fisher 分布の集中度が海底での徘徊で大きい（集中度が弱く分散が大きい）ことは、徘徊と呼ぶにふさわしいジグザグした動きと対応しているように思える。

ただし、上は 1 秒スケールで得られたデータを 3 秒スケールに落として（3 つおきに re-sampling して）作ったデータへ適用した場合である。1 秒スケールのままでは、図 5 のように、残差は Fisher 分布から大きく外れた分布となる。

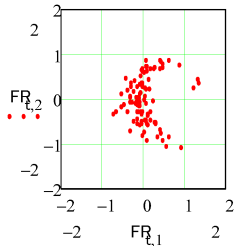


図 5 1 秒スケールで潜水に 3 次元 von Mises process を適用したときの残差。

さらに、軌跡について方向の自己相関

$$C_s = \sum_{t=s+1}^n \overrightarrow{X_t X_{t+1}} \cdot \overrightarrow{X_{t-s} X_{t+1-s}} / (n-s)$$

を計算すると、図 6 のようにモデルが作成する軌跡は実データと外れている様子が伺える。

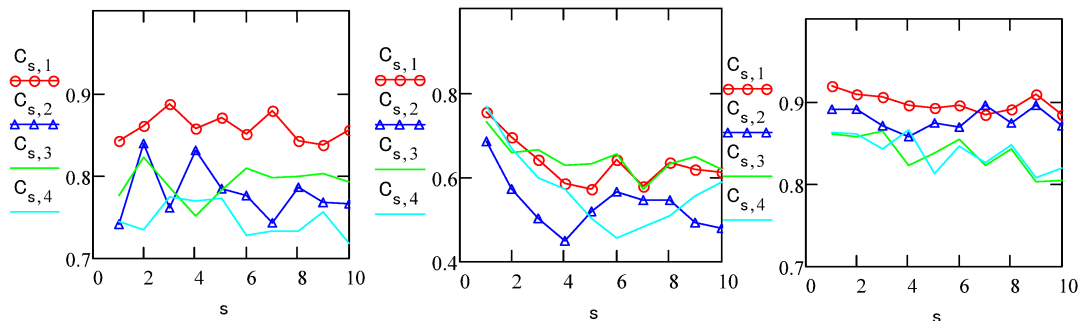


図 6 横軸は自己相関をとる time-lag。赤は実データ、その他の色はシミュレーションで作成した軌跡の方向自己相関。

7. まとめ

2次元の von mises process で定義される方向の自己回帰型モデルは、容易に 3次元へ拡張される。そのモデルが生成する軌跡は様々な形状を描き、実データともよく似たものを見るように見える。しかし、残差や方向自己相関という統計量を用いると、モデルから生成される軌跡は、実は実データとけっこう乖離していることがわかる。この乖離している様子を精査しモデルに足りない面を補っていく事で、より実データを再現するモデルとなっていく。

参考文献

Breckling, J. (1989). The analysis of directional time series: applications to wind speed and direction. Springer-Verlag Berlin.

Kato, S. (2010). A Markov process for circular data. *J. Roy. Stat. Soc. B*, 72, 655–672.

Mardia K. V., P.E. Jupp (2000) Directional Statistics. John Wiley & Sons, Ltd.

Shimatani, I. K., K. Yoda, N. Katsumata, and K. Sato. (2012). Toward the quantification of a conceptual framework for movement ecology using circular statistical modeling. *PloS one* 7:e50309.