

鳥が飛んだGPS軌跡に対する方向モデリング

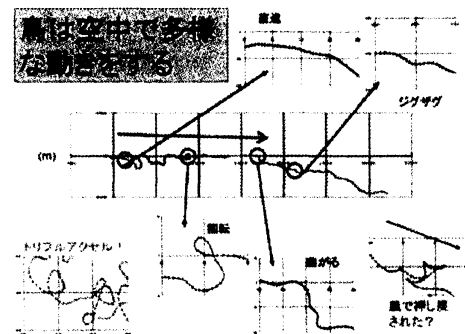
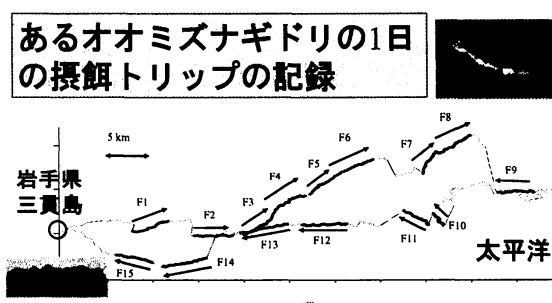
統計数理研究所
島谷健一郎

Ichiro Ken Shimatani
The Institute of Statistical Mathematics

大型の海鳥が強風下の海上でどのように飛んでいるか、実際のところほとんどわかっていない。大型の鳥の飛翔については、風洞実験などの室内実験はほとんど不可能であり、科学的な解明はほとんどされてきていない。しかし、近年の動物装着型小型データロガーの開発により、とりあえず「どのように」飛んでいたかという情報は得られるようになった。GPS軌跡から、実際にどこへ飛んでいったかがわかり、その間の速度も求められる。アホウドリなど滑空する海鳥がどのくらいの頻度でどのような周波数で羽ばたきをはさんでいるかも、加速度ロガーから推定できる。そして、こうした時系列データに統計モデリングを適用することで、さらに多くの新しい発見がある。

ここでは、0.5秒間隔で記録された海鳥の詳細なGPS移動軌跡データに対して、最近試みた結果を報告する。羽ばたきや滑空に伴う流体力学的分析には程遠いレベルであるが、案外と近い将来、海鳥が海上で作る流れについてもデータが取られ分析される時代が訪れるように予見している。

下の左の図は、岩手県釜石市の沿岸にある三貫島で繁殖する海鳥・オオミズナギドリに GPS を装着して得られた移動軌跡データで、右は、その中の一部を取り出したものである。鳥は、まっすぐに移動したり、ジグザグに飛んだり、曲がったり回転したり、多様な飛び方をしている様子が伺える。



軌跡データは、単にどこに行っていたという情報に留まらず、移動パターンに関

する様々な情報を含んでいるはずで、その定量的分析を当面の研究目的に置く。

動物の動きを解析する時に基本となるのがランダムウォークである。このモデルの元では、鳥は基本的に前と同じ方向に飛ぼうとし、若干、誤差でずれるとする。これまで GPS データの解析に関して数多くの論文が発表されているが、ほとんどがランダムウォークを基本としている。

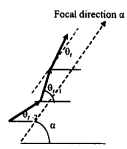
ところが、鳥の実際の軌跡にはランダムな動きというのはほとんど無く、餌場などの決まった方向を目指して飛んでいくというパターンの方が圧倒的に多く見られる。そこで、そういった方向性を持つ動きを表すモデルとして、Nams (2006)は

$$\theta_t = (1-w)(\alpha - \theta_{t-1}) + \theta_{t-1} + e_t$$

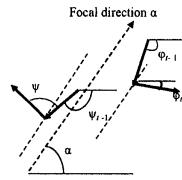
というモデルを提唱した。このモデルでは、次に向かう方向を 0 から 1 までで連続的にパラメトライズする。0 の場合だと目標方向 α へ向かい、1 だと前と同じ方向 θ_{t-1} へ向かう (ランダムウォーク)。中間の 0.5 などでは中間的な方向へ向かう。そこに確率論的な誤差が入る。

目的方向(focal direction)を持つ動きのモデルの例

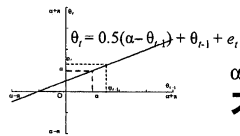
$$\theta_t = (1-w)(\alpha - \theta_{t-1}) + \theta_{t-1} + e_t$$



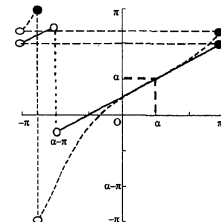
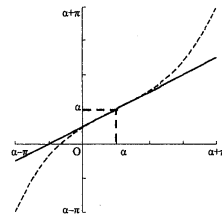
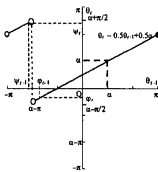
いいモデルに見えるが数学的にも動物学的にも問題がある



グラフを曲げて角度 (円周) から角度 (円周) への滑らかな関数に直す必要性



α の反対方向で不連続



この式で目的方向を有する動きは解決できるように思えるが、実際のところ、ここには数学的にも動物学的にも問題が含まれている。その最たる欠陥は、目的方向と反対方向を境に、曲がる向きが 180 度近く急転する点、即ち数学として不連続である。何らかの形で、滑らかな関数によるモデルに改良する必要がある。

この解決策は Downs & Mardia (2002)及び Kato (2010)により既に提唱されており、左下の様な数式を用いる。そこに、グラフの例も示す。Nams のモデルと同じように、 w が 0 だと目標方向 $\alpha + \beta$ へ進もうとし、1 だと前と同じ方向 θ_{t-1} に進む。そして、1 から 0 になるにつれて目標方向へ向おうという傾向が強くなる。 w が負の値になると、目標方向を通り越した方向転換をする。

数式は複雑に見えるが、右下に述べるように、リンク関数を用いた一般化線形モデルの枠組みのものであり、さらにメービウス変換を実数で表示したものであることも知られている(Kato et al. 2008)。

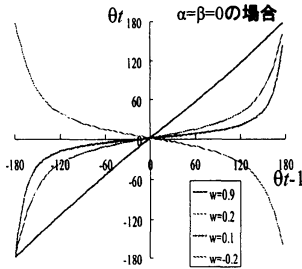
このモデルでは、変換後の方向は一方向 $\alpha + \beta$ に集中する。そのため、ある目標方向に向う軌跡が描け（次々ページの軌跡の例の図参照）、まさしく目的方向を持って進んだ動物の軌跡データに適したモデルとなる。

Kato (2010) の円周自己回帰モデル

$$\theta_t = \alpha + \beta + 2\arctan\{w \tan((\theta_{t-1} - \alpha)/2)\} + e_t$$

w: 回帰係数
 α : 基点
 β : 回転(切片)

e_t : 円周上の確率分布に従う誤差項



線形回帰の素直な拡張になっている

0. 線形回帰 $Y = wX + e_t$, $e_t \sim \text{normal}(0, \sigma^2)$
1. 角度は $(-\pi, \pi)$ なのでリンク関数 $\tan(x/2)$ を入れる $\tan(Y/2) = w \tan(X/2)$
2. 変形してからランダム変動を円周の確率分布に変える $Y = 2\arctan\{w \tan(X/2)\} + e_t$
3. 基点 α を追加 (円周では0が起点とは限らない) $Y = \alpha + 2\arctan\{w \tan((X - \alpha)/2)\} + e_t$

円周を複素数表示: メービウス変換: ごく自然

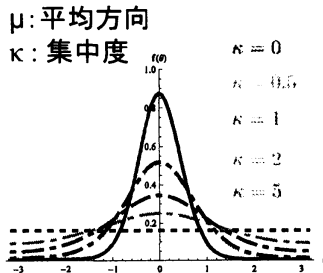
誤差項に用いる確率分布も、円周上の確率分布として $-\pi$ と π で滑らかに繋がっている必要がある。実数上の正規分布は使えないため、von-Mises 分布や、コーシー分布をぐるっと円周に巻きつけた巻き付けコーシー分布などが使われる。どちらも、ひとつのパラメータを変えると、ちょうど正規分布の分散のように様々な集中度の確率分布を作る。

円周の確率分布

von Mises分布

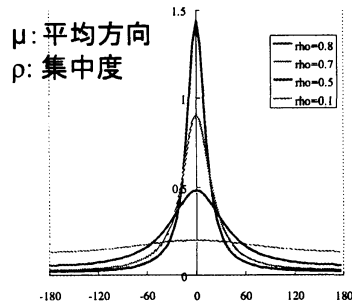
$$f_{VM}(y) = \exp(\kappa \cos(y - \mu)) / 2\pi I_0(\kappa)$$

$I_0(\kappa)$: 指数0の第1種修正ベッセル函数



Wrapped Cauchy 分布

$$f(x) = \frac{1}{2\pi} \frac{1 - \rho^2}{1 + \rho^2 - 2\rho \cos(x - \mu)}$$



このモデルに従い、長さ1のステップで軌跡を描いてみる。左下の図の右のグラフは、それぞれの条件で50ステップをおこなった結果である（あまりに進みがノロかったもののみ100ステップ）。

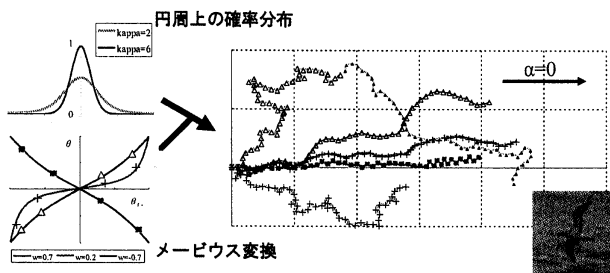
確率分布の密度函数のグラフが尖がっていて誤差が小さいモデルでは、時々エラーはあるものの、目標方向 α に向かって比較的まっすぐ進む。一方、集中のさせ方が弱い (w が1に近い) と、何かの拍子に間違った場合にはしばらく進んでから元

の方向 α に戻る。つまり、 w は鳥が目標方向へ向かおうという意志の強さを反映する。確率分布の集中度が弱い (κ が小さい) と頻繁にエラーがあり、すぐに正しい方向に戻るが、また間違えるといったように、酔っ払いの歩き方に近いものとなる。つまり、 κ というパラメータの値は、行くべき方向へ向かう実行力の強さのようなものを表す。

集中させ方も弱く、しかも確率的にもよく間違える場合には、ふらふらした軌跡になる。

w の値が負になると、毎回、逆方向に向かうようになるので、ジグザグの動きになる。

動物らしい軌跡を描ける



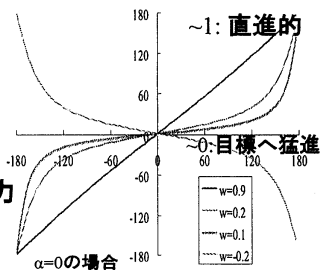
動物行動軌跡では、回帰式の集中は好ましい性質

動物の軌跡データでの解釈

$$\theta_t = \alpha + 2 \arctan \{ w \tan((\theta_{t-1} - \alpha) / 2) \} + e_t$$

- α : 目標方向
- w : 目標へ向かう意志の強さ

e_t の分散 (集中度)
: 目標へ向う実行力



このように、モデルのパラメータ (α, w, κ) は、ある地点において行きたい方向 (α) や行きたい気持ちの強さ (w)、またある方向に行く実行力 (κ) など、それぞれ動物の意志を反映したような解釈を伴う。

実際のデータを適用する時には、このモデルでは確率分布を当てはめるだけなので、容易に対数尤度式を書き下せ、パソコンの計算ソフトで数値的に最適化できる。

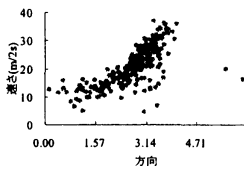
GPS 軌跡データの場合、方向についてのモデルのほかに、速さのモデルも必要である。風向きによって速さは違い、風下だったら移動は早くなり、風上だったら遅くなる。そこで、進んだ方向に応じた速さの違いを考慮するため、簡単な重回帰を用いて「その時の方向」と「前の時点での速さ」でその時の速さを決めるモデルを作り、上記の角度モデルと併せることで軌跡が描けるモデルとなる。

もう一つの問題は、タイムステップに関するものである。 t と $t-1$ の差をいくつにするか。GPS のデータは 0.5 秒毎に取れていても、鳥は方向を短時間の内に調整することはできない。それでは、一体どれくらいの時間間隔で移動方向を調整しているかも、このモデルの枠組みで推定できないだろうか。

一般に、モデルの中では、誤差分布に独立性を仮定し、「独立同分布に従う」

Identically and Independently Distribute (i.i.d.) とする。そこで、データをリサンプリングして、i.i.d.を仮定してよくなるよう、1秒から順にタイムステップを長くしてみた。すると、大体3秒以上で誤差項は独立になった。逆に言うと、3秒ぐらいごとにオオミズナギドリの意識が働き、方向を調整しているのかもしれない、ということである。

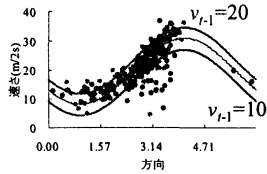
進む速さのモデル



速さ v_t は方向 θ_t と前の速度 v_{t-1} に依存と仮定

$$v_t = a_1 v_{t-1} + a_2 \cos(\theta_t + a_3) + a_4 + \eta_t$$

AICで変数選択



軌跡作成

$$\begin{cases} x_{1t} = v_t \cos \theta_t + x_{1,t-1} \\ x_{2t} = v_t \sin \theta_t + x_{2,t-1} \end{cases}$$

時間間隔 (time unit)

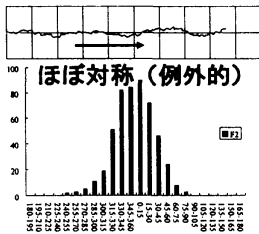
短い時間では方向調整できない

- ・ データは0.5秒ごと
- ・ 1, 2, ..., 10秒ごとにデータを減らしてそれぞれ最尤推定
- ・ 期待値と実測の差に相関のない (e_t にi.i.d.を仮定してよい) 間隔を選ぶ

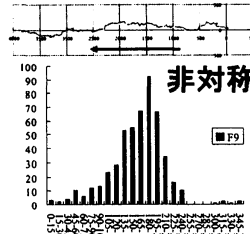
：鳥が動ける最小のtime-scaleをデータとモデルから推定

ところが、実際に岩手県三貫島で繁殖するオオミズナギドリのデータにモデルを当てはめてみると、驚くほどたいいていの場合、当てはまっていなかった。原因は、飛翔方向の非対称性にある。即ち、飛んでいる最中の1秒毎の移動方向を調べると、ほとんどの飛翔でその分布が非対称になっていた。その要因は、風であろう。

しかし、このモデルは大半の軌跡で当てはまらなかった

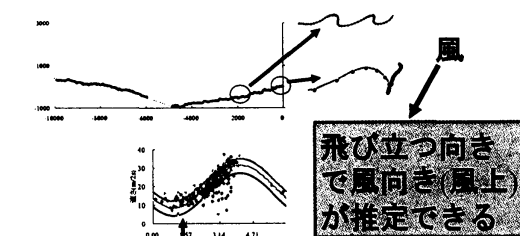


見た目に同じ軌跡が異なる性質を秘めている



1秒ごとの飛んだ方向の分布の例

非対称になる要因のひとつは風



飛び立つ向きは最も低い方向と一致

上のモデルでは対称な確率分布を用いているので、飛んだ方向の分布が対称な軌跡しか描けない。そのため、シミュレーションで描いた軌跡の方向の分布は実データと合致しないし、観察されたゴール地点まで辿り着けない軌跡も多い。

さらに、回帰曲線を進行方向の散布図に入れてみると、目的方向付近とそこから離れた方向とで、分散が違う場合も多い (非対称かつ等分散でない)。つまり、実際の風の中で飛ぶ鳥では、von-Mises 分布などの対称な確率分布を用いた先ほどのモデルは使えないということがわかった。

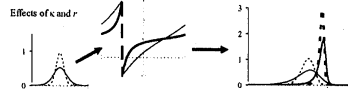
そこで、Kato-Jones の非対称分布を用いることにした。これは、先ほども出てきたメービウス変換を使って、ある方向 μ に集中していた von-Mises 分布を別な $\mu + v$ へ方向捻じ曲げたものである。但し、捻じ曲げにも限界があり、全てのデータが平行移動するわけではなく、 μ にテールを引く分布になる。

この非対称分布は、鳥が風に流された様子を良く表現してくれる。さらに、このモデルの値 r から、風の影響を評価する事もできる。

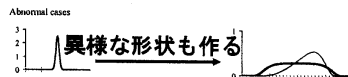
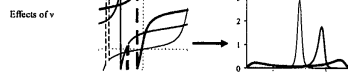
Kato-Jones分布の適用で解決：
平均 μ のvon Mises分布をメービウス変換
 $\mu + v + 2 \tan^{-1} \left(\frac{1-r}{1+r} \tan \left(\frac{\bar{x} - (\mu + v)}{2} \right) \right)$ **で $\mu + v$ へ集中させる**

(A) von Mises distributions (B) Transformations (13) (C) Kato-Jones distributions

集中は不完全
 μ へtailが残る



$\alpha + 2 \tan^{-1} \{ w \tan((\theta_{t-1} - \alpha)/2) \}$ へ進みたい
 のに風で流された
 動きにピッタリ



Kato-Jones分布の確率密度函数

$$f_{KJ}(y; \mu, \kappa, r, v) = \frac{1-r^2}{2\pi_0(\kappa)} \exp \left(\frac{\kappa \{ \xi \cos(y-\eta) - 2r \cos v \}}{1+r^2-2r \cos(y-(\mu+v))} \right) \frac{1}{1+r^2-2r \cos(y-(\mu+v))}$$

$$\xi = \sqrt{r^4 + 2r^2 \cos(2v) + 1}, \eta = \mu + \tan^{-1}(r^2 \sin(2v) / (r^2 \cos(2v) + 1)), -\pi \leq \mu, v < \pi, \kappa > 0$$

$$0 \leq r < 1.$$

尤度式は

$$L(\alpha, w, \kappa, r, \lambda) = \prod_{i=1}^{n-1} f_{KJ}(\Theta_i; m(\Theta_{i-1}), \kappa, r, \lambda - m(\Theta_{i-1}))$$

$$m(\Theta_{i-1}; \alpha, w) = \alpha + 2 \tan^{-1} \{ w \tan((\Theta_{i-1} - \alpha)/2) \}$$

本来行こうと思っていた方向 $m(\Theta_{i-1})$ が λ への風により $\lambda - m(\Theta_{i-1})$ 曲がって λ へ集中

結局、モデルとしては5つのパラメータ ($\alpha, w, \kappa, \mu, \lambda$) を有し (風向き λ は、速さが最大となる方向としてもよい)、下のような解釈ができる。なお、誤差分布としては変換前の von Mises の段階で i.i.d. を仮定するので、観察された方向をメービウス逆変換した方向を用いて確かめることになる。また、パラメータに何らかの制限を入れることで、様々な仮説の検証も可能となる。

time unit: 観察値 Θ_i をメービウス逆変換した角度と $m(\Theta_{i-1})$ の差に自己相関がない: motion capacity の time scale

α : 目標方向
 w : 目標へ向かう意志の強さ
 r : 風の影響の強さ
 κ : 目標へ向かう実行力

以下のような制約を入れたモデルもAICで評価

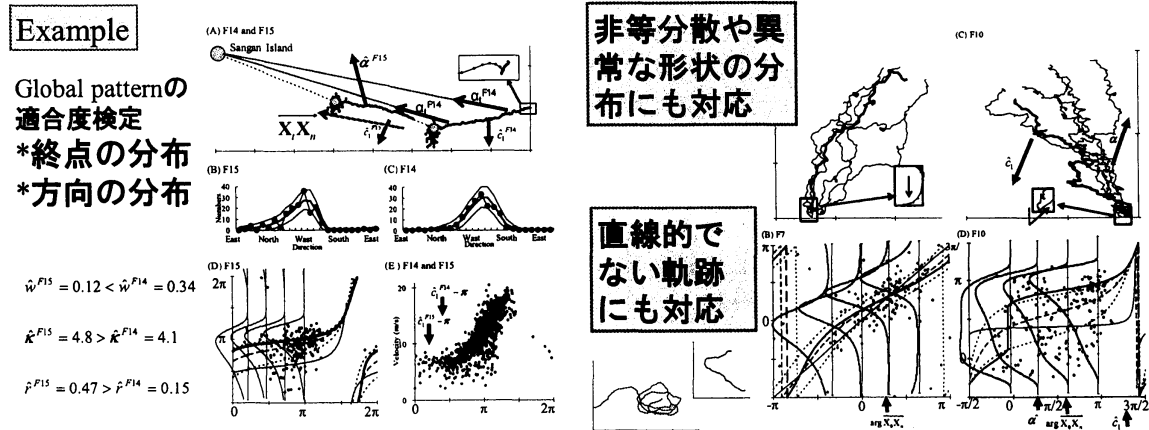
λ =速さ最小の方向: 本当に風下に流されたか?

$\alpha = \arg X_t X_{t-1}$ 目標地点(営巣している島)に向きたい?

$\alpha = \arg X_t X_{t-2}$ 計画通りに飛べたか?

Kato-Jones の非対称分布を用いたモデルは、実際のデータに極めて良好な当てはまりを示した。ここでは3つの事例のみ示す。島に帰る最後の飛翔では、前半は鳥は島に帰るつもりで飛んでいたが、意志が弱かったために (w が 1 に近い、 κ が小さい) 風に流されて島から近づけなかった。それで途中から意志を強く持って飛んだ。別な例では、分散は対称だが非等分散なので Kato-Jones 分布が必要になったり、

大きな非対称分散をしても Kato-Jones 分布で適合させられている様子が伺えた。



動物に小型記録計を装着して行動を記録するバイオロギングという分野では、ベースになる良いモデリングがなく、それがデータ解析の最大の障害だった。それが、GPS 軌跡に対しては、出発点になるものができたと言える。

優れた基本モデルがあると、それで説明できない部分は特異な現象に対応し、鳥が餌を見つけた瞬間などの発見にもつながるはずである。Kato の円周自己回帰に Kato-Jones 分布を誤差分散として用いると、飛ぶ鳥の基本モデルとして申し分ないものになっているように思う。

<p>結び</p> <ul style="list-style-type: none"> ・ GPS軌跡データの基本モデルがようやく誕生 (AICで定性的に分類、最尤推定値定量的で評価) 1) 様々な実データによく当てはまる 2) パラメータに生物学的解釈 3) 数学的に自然、PCで計算可能 <p>・ 海上の強風下で鳥はどう飛んでいるかデータとモデリングで推測</p>	<p>このモデルで説明できない部分がより高度な意志が働いている動物学的にも重要な行動</p> <ol style="list-style-type: none"> 1. ランダムウォークの数理などと比べて、軌跡データの統計モデルは著しく未発達 2. 線を引くだけの回帰よりはメカニズムに基づくモデル 3. すぐれた確率分布が不可欠
--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

参考文献

Downs, T. D., and K. V. Mardia. 2002. Circular regression. *Biometrika* 89:683-697.

Kato, S., and M. C. Jones. 2010. A family of distributions on the circle with links to, and applications arising from, Möbius transformation. *Journal of the*

American Statistical Association 102:249-262.

Kato, S., K. Shimizu, and G. S. Shieh. 2008. A circular-circular regression model. *Statistica Sinica* 18:633-645.

Kato, S. 2010. A Markov process for circular data. *Journal of the Royal Statistical Society B* 72:655-672.

Mardia, K. V. and P. E. Jupp. 2000. *Directional Statistics*, John Wiley & Sons, Ltd, England.

Nams, V. O. 2006. Detecting oriented movement of animals. *Animal Behaviour* 72:1197-1203.