第 10 回横幹連合コンファレンス
2019.11.30-12.1 長岡技術科学大学

# さまざまな研究パラダイムをつなぐ情報幾何

江口 真透 （統計数理研究所）

**Information geometry combining various research paradigms**

S. Eguchi (Institute of Statistical Mathematics)

**Abstract**– Information Geometry has been advocated by Shun-ichi Amari in early 1980s, which provides a geometric insight for understanding statistical ideas such as information, sufficiency and efficiency. Furthermore, it is closely connected with almost all of areas in mathematical sciences including information, statistical, physical, biological and brain sciences. One of the most characteristic of information geometry is in a dualistic pair of e-connection and m-connection in the space of all probability density functions, which can be viewed as a Riemannian space with a metric tensor derived by the Fisher information matrix. Surprisingly, the essential theorem can reduce to the Pythagorean theorem developed in the ancient Greek era, which provides a view for the interplay between a statistical model and estimation expanding a Pythagorean foliation in the probability density function space. Finally, we have a review for the present address and future directions in information geometry.

**Index terms**– e-connection, information metric, m-connection, statistical model, statistical estimation

## 1 推薦対象および推薦理由の概要

情報幾何 (Information Geometry) をコトつくりの至宝に相応しい対象として推薦する．甘利俊一氏は大学院生時代から情報空間の幾何的考察を開始 [5, 7] していたが 1980 年代初めに情報幾何の一般的な枠組みの構築に成功した [1]．この枠組みでは確率密度関数の空間に情報計量が導入され，さらに m-接続と e-接続と呼ばれる線形接続が導入された．この 2 つの線形接続は情報計量によるリーマン接続に関して双対的であることが示された．このように確率密度関数の空間に対して双対リーマン空間としての枠組みが構築された．更にアファイン微分幾何として深化が図られてきた [14, 16]．

この枠組みによって統計学の基本的な考えである統計モデルの情報量，統計量の十分性，推定量の有効性の幾何的理解がもたらされた．特に e-接続の平坦性は統計モデルの最適性を，m-接続の平坦性は推定の最適性を曲率テンソルを使い定量的に計ることを可能にした．これによって統計学に限らず，ランダムネスを扱う様々な研究分野，**情報理論，信号処理，量子物理，人工知能，機械学習**などに大きな影響を与えている．

このように情報幾何が提唱されて以来，多くの日本の研究者に影響を与え，日本発の研究として情報幾何の研究が盛んになり，同時に世界的な展開が成されている．最近では最適輸送問題，深層学習などの最先端の研究にも深く関わっており，2018 年に Springer から **Information Geometry**[12] が刊行され情報幾何の精神による新たな発想から実問題の解決から数理的な新概念の構築に渡って，更なる挑戦が継続されている．このことから情報幾何はコトつくりの至宝として十分な価値を持つと考えられる．

## 2 情報幾何が作り出す世界

確率密度関数全体の空間 $\mathcal{P}$ に統計モデル $M = \{p_\theta(x) \in \mathcal{P} : \theta \in \Theta\}$ を考えよう．パラメータ $\theta$ のフィッシャーの情報行列

$$\mathrm{I}(\theta) = \mathrm{E}_{p_\theta}\Big[\frac{\partial \log p_\theta(X)}{\partial\theta}\frac{\partial \log p_\theta(X)}{\partial\theta^\top}\Big]$$

をリーマン計量として定めたとき $M$ はリーマン空間になっていることが示された [21]．ここで $E_{p_\theta}$ は確率密度 $p_\theta$ に関する期待値を表す．パラメータ $\theta$ の不偏推定量が存在するとき $\mathrm{I}(\theta)^{-1}$ が不偏推定量の分散行列の下限（クラメール・ラオ下限）を与えることからリーマン計量として定めることが正当化される．このように定められたリーマン計量のことを情報計量と呼ぶ．以来，幾つかの研究によって統計モデルの幾何的研究が進められた．

甘利は 1982 年にリーマン空間の中で双対な線形接続の **e-接続**と **m-接続**を導入し，それぞれが統計モデルと推定に基本的な尺度を与えることを示す画期的な成果を発表した [1]．その理論の中で本質的な定理は以下に紹介されるよう，驚くべきことに古代ギリシャ数学の**ピタゴラスの定理**にまで遡る [17, 4]．

確率密度関数空間 $\mathcal{P}$ に 3 角形を考えたとき，次のピタゴラスの 3 平方定理が成立する：
$\mathcal{P}$ の 3 点 $p, q, r$ からなる 3 角形 $\Delta(p, q, r)$ を考えよう．このとき，頂点 $p$ と $q$ を結ぶ m-測地線 $\{p_t(x) : 0 \le t \le 1\}$ と頂点 $q$ と $r$ を結ぶ e-測地線 $\{r_t(x) : 0 \le t \le 1\}$ を

$$p_t(x) = (1-t)p(x) + tq(x)$$
$$r_t(x) = c_t \exp\{(1-t)\log r(x) + t\log q(x)\}$$

と定める ($c_t$ は規格化定数)． このとき，2 つの測地線

が点 $q$ で情報計量の意味で直交するとき，

$$D(p,q)+D(q,r)=D(p,r) \tag{1}$$

が成り立つ (Fig. 1 参照)．ただし，辺の長さの 2 乗はつぎの KL ダイバージェンスによって取られる：

$$D(p,q)=\int p(x)\{\log p(x)-\log q(x)\}dx$$

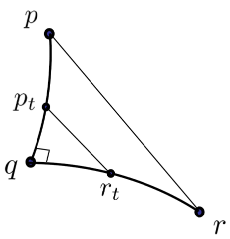


Fig. 1: Pythagoras theorem

この定理の証明を詳しくみると，任意の $t\in[0,1]$ で

$$D(p_t,q)=D(p_t,q)+D(q,r_t)$$

が成立することが分かる．このようにユークリッド幾何のピタゴラス定理は直線の代わりに e-測地線と m-測地線を取ると確率密度空間 $\mathcal{P}$ で成立することから，$\mathcal{P}$ は**双対ヒルベルト空間**と考えることもできる．また，ガウスの最小二 2 乗法の双対的拡張と見れ，最尤推定の満たす基本性質となっている．AIC との関連も深い．

上で見てきたように (1) を満たすピタゴラス定理の情報幾何的な考察をしよう．そのため，指数モデル

$$M^{\mathrm{e}}=\{p^{\mathrm{e}}_{\theta}(x):=\exp\{\theta^{\top}s(x)-\psi(\theta)\}:\theta\in\Theta\}$$

を考える．ここで $\psi(\theta)$ はキュムラントと呼ばれ

$$\psi(\theta)=\log\int\exp\{\theta^{\top}s(x)\}dx$$

を表し，$\Theta=\{\theta\in\mathbb{R}^p:\psi(\theta)<\infty\}$ とする．この時，$s(X)$ を十分統計量と呼び，平均一致空間

$$\mathcal{N}^{\mathrm{m}}(p)=\big\{q\in\mathcal{P}:E_q\{s(X)\}=E_p\{s(X)\}\big\}$$

を考えよう．指数モデル $M^{\mathrm{e}}$ の任意の点 $p,q$ を結ぶ e-測地線は $M^{\mathrm{e}}$ に含まれ，平均一致空間 $\mathcal{N}^{\mathrm{m}}(p)$ の任意の点 $q,r$ を結ぶ m-測地線は $\mathcal{N}_{\mathrm{m}}$ に含まれる，すなわち，それぞれの接続に関して全測地的である．このように，

(i) $\quad p\neq q$ in $M\Rightarrow\mathcal{N}^{\mathrm{m}}(p)\cap\mathcal{N}^{\mathrm{m}}(p)=\emptyset$

(ii) $\quad \mathcal{P}=\bigcup_{p\in M}\mathcal{N}^{\mathrm{m}}(p)$

を満たす葉層が連想されることが分かる．さらに $M^{\mathrm{e}}$ の点 $q$ をかってに固定すると，任意の $p\in\mathcal{N}^{\mathrm{m}}(q)$ と任意の $r\in M^{\mathrm{e}}$ に対してピタゴラス定理 (1) が成立する．従って，

$$D(p,q)=\min_{r\in M^{\mathrm{e}}}D(p,r)$$

が言えるので $\mathcal{N}^{\mathrm{m}}(q)$ から $M^{\mathrm{e}}$ への KL ダイバージェンス最小化は全て 1 点 $q$ で達成される．KL ダイバージェンス最小化は尤度最大化と等価なので $\mathcal{N}^{\mathrm{m}}(q)$ は最尤推定葉と呼ばれる．このように十分統計量 $s(X)$ が生成する指数モデル $M^{\mathrm{e}}$ を茎とし，指数モデル $M^{\mathrm{e}}$ の各点 $p$ に最尤推定葉 $\mathcal{N}^{\mathrm{m}}(p)$ が連想され，$p$ を $M^{\mathrm{e}}$ 上で動かすと平行な葉 $\mathcal{N}^{\mathrm{m}}(p)$ が連なる**ピタゴラス葉層構造**ができる（Fig. 2 参照）．

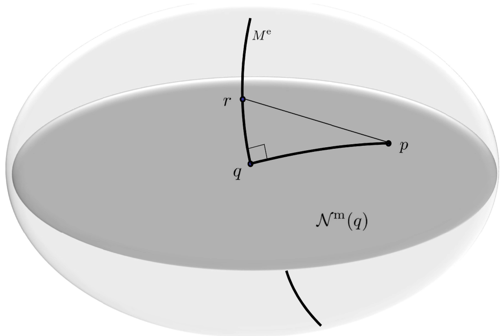


Fig. 2: Pythagoras foliation

以上の考察で指数モデル $M^{\mathrm{e}}$ と最尤推定葉 $\mathcal{N}^{\mathrm{m}}$ の相補的関係が互いに双対な e-接続と m-接続によって明らかにされた．これは統計モデリングと推定という 2 つの操作は一つの基準での統一はできなく，別々の基準が必要であり，それらが双対関係によって結ばれることが明らかにされて初めて 2 つの操作間の相互作用が記述できることを物語っている．物理学では 4 次元時空間の重力テンソル計量によって光の経路が最短経路を定めるリーマン測地線として与えられ，一般性相対理論の壮大な力学的世界像が構築された．情報幾何では双対な測地線がモデルと推定という相補的な関係を明らかにしてしている．

一般に $\mathcal{P}$ の正則なダイバージェンスは任意のモデル $M$ に制限するとリーマン計量と 1 組の双対接続が連想される [9, 10]．その双対な測地線によってピタゴラス定理が成立することが示せる．ダイバージェンスが対称であるならば一組の双対接続はリーマン接続に帰着される．KL ダイバージェンス $D(p,q)$ を採用すると情報計量が導かれ，$D(p,q)$ の非対称性から e-接続と m-接続が導かれ，その非対称性とネイマン・ピアソン定理と密接な関係が分かる [11]．

## 3 まとめ

情報幾何はリーマン空間での双対接続という意味では確立され，基礎的な枠組みができあがり，国際会議 Information Geometry and its Applications が 4 回開

催され（ペスカラ，2002；東京，2005；ライプニッツ，2010；リブリツェ, 2016）を含む様々な国際会議，研究集会を通して活発な交流からを多くの発展を遂げている．来年も東京で情報幾何学国際会議 2020 が計画されている．情報幾何の今後に解決すべき問題や，未開拓な方向もたくさんあると言える．以下に幾つかの最近の動向と今後の方向についてまとめる．

幾何学の最近の発展に**最適輸送問題**が挙げられる[23, 15]．この問題は 18 世紀のモンジュの問題提起から始まり，確率分布間のワッサースタイン距離による定式化がされた．このように確率分布のクラスを考えることから情報幾何との融合が盛んになる機運が盛り上がっている[20, 19, 24, 22]．

人工知能，機械学習の分野で**深層学習**の再興が起こり，様々な自動化が可能になり人間社会の構造も変化がもたらされている．特に畳み込みニューラルネットワークの学習の高速化と汎用化がリアルタイムの複雑な画像認識を可能にしている．これによって強化学習の高度な計算が実時間で可能になり，従来，人間が高い負担を伴なってやってきた作業や長年の鍛錬が必要な技術も自動化できるものが拡大されている．情報幾何はニューラルネットワークの数理について深い研究[2, 3, 6]があり，刷新された深層学習についても密接な展開が成されつある[22]．最適輸送問題のエントロピー正則化によって深層学習と情報幾何との融合が進む期待が高まっている．また，近年，多様体学習はシェイプ解析など数学的追及と高度な応用が目指される分野に変容しつつある[18, 8]．量子物理の情報幾何による枠組みが与えられ[13]，**量子情報論**への更なる拡張が期待されている．統計物理もタリス・エントロピー，べき分布との関連から研究されているが，近い将来，更なる突破口が開かれることが期待される．

このように情報幾何は統計学の基本原理への貢献から端を発し，様々な数理科学の分野に本質的な関連の糸口が切り開かれてきた．しかし，その可能性を十分に追及され，確立された分野とは言い難い．寧ろ，多くの可能性は未だ十分に開発された状況にないと言える．更なる壮大な構想の下に情報幾何の大きな成果が収穫される日が到来することが待たれてならない．

## 参考文献


1) Amari, S. I. (1982). Differential geometry of curved exponential families-curvatures and information loss. The Annals of Statistics, 357-385.

2) Amari, S. I. (1995). Information geometry of the EM and em algorithms for neural networks. Neural networks, 8(9), 1379-1408.

3) Amari, S. I. (1998). Natural gradient works efficiently in learning. Neural computation, 10(2), 251-276.

4) Amari, S. I., & Nagaoka, H. (2007). Methods of information geometry (Vol. 191). American Mathematical Soc.

5) Amari, S. I. Information Geometry: Historical episodes and recent developments. Information Geometry and its Applications IV June 12-17, 2016, Liblice, Czech Republic

6) Amari, S. I. (2016). Information geometry and its applications (Vol. 194). Berlin: Springer.

7) 甘利俊一. "Information Geometry"（情報幾何学）第 23 回高木レクチャー．京都大学数理解析研究所

8) Dryden, I. L. (2014). Shape analysis. Wiley StatsRef: Statistics Reference Online.

9) Eguchi, S. (1983). Second order efficiency of minimum contrast estimators in a curved exponential family. The Annals of Statistics, 793-803.

10) Eguchi, S. (1992). Geometry of minimum contrast. Hiroshima Mathematical Journal, 22(3), 631-647.

11) Eguchi, S., & Copas, J. (2006). Interpreting Kullback-Leibler divergence with the Neyman-Pearson lemma. Journal of Multivariate Analysis, 97(9), 2034-2040.

12) Information Geometry, 1.1. Springer. (2018)

13) Hayashi, M., & Nagaoka, H. (2003). General formulas for capacity of classical-quantum channels. IEEE Transactions on Information Theory, 49(7), 1753-1768.

14) Kurose, T. (1994). On the divergences of 1-conformally flat statistical manifolds. Tohoku Mathematical Journal, Second Series, 46(3), 427-433.

15) 最適輸送理論とリッチ曲率．桑江一洋, 塩谷隆, 太田慎一．日本数学会, 2017

16) Matsuzoe, H. (1998). On realization of conformally-projectively flat statistical manifolds and the divergences. Hokkaido Mathematical Journal, 27(2), 409-421.

17) H. Nagaoka and S. Amari, Differential geometry of smooth families of probability distributions, Technical Report METR 82-7, Dept. of Math. Eng. and Instr. Phys, Univ. of Tokyo, 1982

18) Ohara, A., Suda, N., & Amari, S. I. (1996). Dualistic differential geometry of positive definite matrices and its applications to related problems. Linear Algebra and its Applications, 247, 31-53.

19) Pal, S., & Wong, T. K. L. (2018). Exponentially concave functions and a new information geometry. The Annals of Probability, 46(2), 1070-1113.

20) Peyre, G., & Cuturi, M. (2019). Computational optimal transport. Foundations and Trends in Machine Learning, 11(5-6), 355-607.

21) Rao, C. R. (1945). Information and the accuracy attainable in the estimation of statistical parameters. Bull. Calcutta Math. Soc. 37, 81-89.

22) Sonoda, S., & Murata, N. (2019). Transport analysis of infinitely deep neural network. The Journal of Machine Learning Research, 20(1), 31-82.

23) Villani, C. (2008). Optimal transport: old and new (Vol. 338). Springer Science & Business Media.

24) Wong, T. K. L. (2018). Logarithmic divergences from optimal transport and Renyi geometry. Information Geometry 1 (1), 39-78.