シリーズ 情報科学における確率モデル 5

Series on Stochastic Models in Informatics and Data Science

# エントロピーの幾何学

田中　勝【著】

コロナ社

# 刊行のことば

われわれを取り巻く環境は，多くの場合，確定的というよりもむしろ不確実性にさらされており，自然科学，人文・社会科学，工学のあらゆる領域において不確実な現象を定量的に取り扱う必然性が生じる。「確率モデル」とは不確実な現象を数理的に記述する手段であり，古くから多くの領域において独自のモデルが考案されてきた経緯がある。情報化社会の成熟期である現在，幅広い裾野をもつ情報科学における多様な分野においてさえも，不確実性下での現象を数理的に記述し，データに基づいた定量的分析を行う必要性が増している。

一言で「確率モデル」といっても，その本質的な意味や粒度は各個別領域ごとに異なっている。統計物理学や数理生物学で現れる確率モデルでは，物理的な現象や実験的観測結果を数理的に記述する過程において不確実性を考慮し，さまざまな現象を説明するための描写をより精緻化することを目指している。一方，統計学やデータサイエンスの文脈で出現する確率モデルは，データ分析技術における数理的な仮定や確率分布関数そのものを表すことが多い。社会科学や工学の領域では，あらかじめモデルの抽象度を規定したうえで，人工物としてのシステムやそれによって派生する複雑な現象をモデルによって表現し，モデルの制御や評価を通じて現実に役立つ知見を導くことが目的となる。

昨今注目を集めている，ビッグデータ解析や人工知能開発の核となる機械学習の分野においても，確率モデルの重要性は十分に認識されていることは周知の通りである。一見して，機械学習技術は，深層学習，強化学習，サポートベクターマシンといったアルゴリズムの違いに基づいた縦串の分類と，自然言語処理，音声・画像認識，ロボット制御などの応用領域の違いによる横串の分類によって特徴づけられる。しかしながら，現実の問題を「モデリング」するためには経験とセンスが必要であるため，既存の手法やアルゴリズムをそのまま

適用するだけでは不十分であることが多い。

本シリーズでは，情報科学分野で必要とされる確率・統計技法に焦点を当て，個別分野ごとに発展してきた確率モデルに関する理論的成果をオムニバス形式で俯瞰することを目指す。各分野固有の理論的な背景を深く理解しながらも，理論展開の主役はあくまでモデリングとアルゴリズムであり，確率論，統計学，最適化理論，学習理論がコア技術に相当する。このように「確率モデル」にスポットライトを当てながら，情報科学の広範な領域を深く概観するシリーズは多く見当たらず，データサイエンス，情報工学，オペレーションズ・リサーチなどの各領域に点在していた成果をモデリングの観点からあらためて整理した内容となっている。

本シリーズを構成する各書目は，おのおのの分野の第一線で活躍する研究者に執筆をお願いしており，初学者を対象とした教科書というよりも，各分野の体系を網羅的に著した専門書の色彩が強い。よって，基本的な数理的技法をマスターしたうえで，各分野における研究の最先端に上り詰めようとする意欲のある研究者や大学院生を読者として想定している。本シリーズの中に，読者の皆さんのアイデアやイマジネーションを掻き立てるような座右の書が含まれていたならば，編者にとっては存外の喜びである。

2018 年 11 月

編集委員長　土肥　正

# ま　え　が　き

情報科学において，データの入出力関係やノイズそのものに対して確率モデルを考えることは，いまや常識となって久しい。さらに，深層学習の発展により，なぜ深層学習がうまく機能するのかという理論的解析やそのアルゴリズムの改良に際してもいまでは確率・統計的な考察は必要不可欠なものとなっている。

ところで，確率モデルを考えるうえで，これまで考えられてきたものはおもに指数型分布族が中心であった。しかし，自然言語処理で前処理として重要な役割を持っている word2vec の高速化のために登場してきた negative sampling の手法では，負例の登場確率が小さなものでもある程度選ばれやすくするために，その確率をべき乗[†1]した後，規格化して得られるエスコート分布を用いることで，さらなる性能の向上が図られている。さらに，通常 generative adversarial network（GAN）の学習は不安定[†2]になりやすいが，安定した学習が可能な Wasserstein GAN[†3]が提案されたことにより，指数型分布族のみならず非指数型分布族[†4]も容易に考察することができるようになってきた。このような状況のなか，確率モデルとしてニューラルネットワークを考えるとき，確率・統計の取扱いについてある程度深く知っておくことが重要になってきている。特に，海外の研究者の書く関連分野の論文等では，測度論的確率論に基づく記述も非常に多い。一方，国内に目を向ければ，むしろ避けられているようでさえある。このような状況を少しでも改善し，さらに確率モデルを記述し，そのモデルについて考

†1　通常，0 と 1 の間の数が選ばれる。例えば，0.75 が選ばれたりする。詳細について興味のある読者は，例えば，巻末の文献[17)] の pp.154–158 を読まれるとよい。

†2　いわゆる，勾配消失問題である。

†3　特徴は，Jensen-Shannon ダイバージェンスを Wasserstein 距離で置き換えるところである。この Wasserstein 距離は，earth mover’s distance と呼ばれることもある。

†4　Wasserstein 幾何学では，非指数型分布族である $q$-正規分布族を導出することも可能である。本書では，それとは異なる方法で $q$-正規分布族を導出する。

察することを容易にするためには，どこかで一度必要な範囲で測度論的確率論に触れる必要がある。そこで本書では，測度論的確率論を，Radon-Nikodým の定理の証明を理解することを目標に紹介することにした。この Radon-Nikodým の定理は，確率密度関数が存在することを保証するものとして見ることもできるが，現代では条件付き確率の存在を保証する定理として，より重要性を増している。他書を参照することなく，本書のみで理解できるように丁寧な式変形と解説を心掛けたが，うまくいっただろうか。

さらに，パラメトリックな確率モデルを考えるときに必要になる母数（分布を特徴付けるパラメータ）の推定に関して必要となる十分統計量の説明も行った。この部分は，文献[18] の pp.1–7 に基づいているが，定義や定理等の書き方および証明については，文献[19] の pp.111–120 を参考にしている。しかし，どちらも初めて数理統計学に接する者には読みにくいため，より証明を詳細に記述することで，できるだけ行間を読まなくても済むように配慮した。

さて，確率モデルを考えるうえで，まず必要になることは確率分布についての知識である。指数型分布族に関しては，甘利 俊一 氏により整理され発展してきた指数型分布族に対する情報幾何学が，その後，多くの研究者によりいまも活発に研究されている。この情報幾何学を用いることで，推定量の有効性や em アルゴリズム，ターボ符号，Boltzmann machine やニューラルネットワークの情報幾何学など多くの展開がなされている。そこで重要な役割を演じるのは拡張された一般化 Pythagoras の定理である。この定理の威力により，さまざまな問題が幾何学的に理解できるようになり，直感的な理解が可能となっている。すなわち，$e$-射影と $m$-射影による交互正射影である。

一方で，非指数型分布族については，いまだ定番となった幾何学は存在していない。そこで本書では，測度空間に特別な平行移動を導入することでアファイン空間を構成し，そのうえで幾何学を展開することにより指数型分布族と非指数型分布族を同時に取り扱うことができるような枠組みを提供することで，確率分布族についての普遍的な性質をとらえるための舞台を紹介することにした。この型の情報幾何学を甘利らの情報幾何学と区別するために $\tau$-情報幾何学

と呼んでいるが，読者が甘利らの情報幾何学に関する書籍であると勘違いすることを恐れたため，本書のタイトルは“エントロピーの幾何学†”とした。本書で紹介する $\tau$-情報幾何学は，甘利らの情報幾何学とは，$\tau = 1$ のときに $\alpha = 1$，$\tau = 0$ のときに $\alpha = -1$ に対応している。しかし，パラメータ $\tau$ と $\alpha$ とでは，その役割がまったく異なることに注意する必要がある。パラメータ $\tau$ は確率分布族を決定し，さらにエントロピーやダイバージェンスも決定してしまう。これに対してパラメータ $\alpha$ は双対接続を表現するために導入され，$\alpha$-ダイバージェンスのパラメータ $\alpha$ とは無関係に設定することができる。最近では，甘利らの情報幾何学は，江口 真透 氏により提案されたものであるが，まず凸関数を一つ与え，それに基づいてダイバージェンスを構成し，このダイバージェンスからさまざまな幾何学的量を導出していくスタイルが主流である。

最後に，本書の測度論的確率論の部分と十分統計量の部分を丁寧に読み貴重なコメントをお寄せいただいた福岡大学の天羽 隆史 氏に感謝いたします。もし，その部分に誤り等があればもちろん筆者の責任であることはいうまでもないことである。また，本書の執筆を勧めていただいた広島大学の栗田 多喜夫 氏に感謝いたします。最後に，コロナ社には，筆者の遅筆にも関わらず辛抱強く待っていただいたことに感謝いたします。

2019 年 3 月

田中　勝

† 甘利らの情報幾何学では，エントロピーはポテンシャル関数と呼ばれるキュミュラント母関数の双対ポテンシャルと異符号の関係にあり，指数型分布族の幾何学を特徴付ける量になっている。一方，$\tau$-情報幾何学でのエントロピーは，べき型に拡張された一般化エントロピーになっており，そのままでは双対ポテンシャルとしての役割を持つことはできず，共形変換を施した後に双対ポテンシャルとしての役割を持つことが可能となる。べき型に拡張された一般化エントロピーをある拘束条件のもとで最大にする確率分布として指数型分布族をべき型に拡張した確率分布族が得られる。この確率分布族は，もちろん非指数型分布族になっている。

# 目　　次

# 1 本書の構成

ここに，本書の構成を示す。何を学んでいるのか不安になったときは，もう一度ここを読んで確認してほしい。

第 2 章では，測度論的確率論の基本的事項を von Neumann による Radon-Nikodým の定理の証明[8)]†1 を理解することを目標にして紹介した。まずは，さいころの例を用いて基本的用語のイメージを具体的にとらえ，その後，より抽象的で一般的な解説へ進むことにした。そこでは積分の定義も非負単関数による下からの近似の極限として与えられる。これは，Lebesgue 積分への第一歩である†2。

第 3 章では，測度空間に Radon-Nikodým の定理を用いることで，ある種の平行移動を導入する。これにより測度空間は一つのパラメータによって決定されるアファイン構造，すなわち $\tau$-アファイン構造を持つことになる。この構造に基づいて，非指数型分布族を指数型分布族と同様に取り扱うことができる舞台が自然に導入されることになる。しかし，この $\tau$-アファイン構造は平行移動により測度の大きさを保存しない。そこで，そのことを積極的にとらえて測度の大きさ方向の次元を一つ追加して考えることにし，スケール変換も同時に考えることにした。このようにして導入された非指数型分布族の取扱いは，十分統計量の拡張として見ることもできる。また，確率変数の空間を $r$ 次の多項式空間にとることで，$r$ 次元の自然座標系が導入される。これは確率分布族の $r$-

†1　肩付き数字は巻末の引用・参考文献を表す。
†2　この定義を Lebesgue 積分の定義だと思っていてもユーザとしては特に困ることはない。

ジェット空間を考えていることに対応している。また，$\tau$-対数尤度を自然座標で微分して得られるスコア関数の型を見ると，$\tau = s$ と $\tau = 1 - s$ を組みにして取り扱うと便利なことがすぐにわかる。そこで，*Body* と *Soul* が導入される。

第 4 章では，$\tau$-アファイン構造は平行移動により，測度の大きさを保存しないので，そのことも考慮し平行移動の仕方に応じて座標がどのように更新されるのかを示した。つまり，始点から終点に至る経路順序により座標の更新の仕方が決まっており，それがどのような規則に従っているのかを具体的に示す。

第 5 章では，*Body* 世界の量と *Soul* 世界の量を組み合わせることで現実世界（*Real*）の量を構成する縮約という操作を定義する。これにより，まずは Fisher 計量を導く。その結果，不定計量になっていることがわかる。この計量から得られるアファイン双対接続は，Koszul 接続になっていることも確かめられる。このような計量を用いた応用例として，Cramér-Rao の不等式の証明を与えることにした。

第 6 章では，くり込み† を用いてエントロピーを定義する。このエントロピーは非加法性を持っているが，べき型の対数関数の性質として 2 種類の恒等式が成り立つために優加法性になる場合と劣加法性になる場合が起こり得る。共形エントロピーとの相性がよいのは，劣加法性のほうであるため，本書では劣加法性の場合を考えていく。このことについては，第 9 章で詳しく見ていくことになる。また，ここで定義されたエントロピーと Boltzmann-Shannon エントロピー，Rényi エントロピー[11)]，Havrda-Charvát エントロピー（Tsallis エントロピー[10)] と呼ばれることもある）との関係についても明確にされる。この後，ダイバージェンスを，平行移動後に得られる測度を平行移動量で一次近似した際の誤差として定義する。この定義は Bregman ダイバージェンスにもなっている。ここで定義されるダイバージェンスの双対ダイバージェンスも定義されるが，これらは甘利らの情報幾何学での $\alpha$-ダイバージェンスと類似の性質を

† 発散項をあらかじめ取り除いておくことで，意味のある有限値を得る手段である。現在では，K.G. Wilson によりくり込み群（renormalization group）として体系化されている。

持っている。そのため，一般化 Pythagoras の定理[1)~7)]も成立することを確認することができる。このことにより，応用に関しては $\tau$-情報幾何学は甘利らの情報幾何学と同様な扱いができることになる。

第 7 章と第 8 章では，具体例としておもに $q$-正規分布について考えていく。まず，第 7 章で $q$-正規分布について紹介した後，第 8 章で，一般的な状況での $\tau$-変換について考察した後，$q$-正規分布に関する $\tau$-変換[12)] について議論する。これにより，いわゆるエスコート分布が分散をスケール変換した別の $q$-正規分布になっていることが示される。つまり，確率分布からエスコート分布を生成し，それを用いて期待値をとることは，別の確率分布を用いて期待値をとることに対応しているのである。その際，$q$-正規分布の場合では，$q > 1$ のときには，より裾が軽い確率分布に変換されることになり，$q < 1$ のときには，より裾が重い確率分布に変換されることになることがわかる。また，正規分布は $\tau$-変換のもとで不変な確率分布になっている。このような $\tau$-変換は，$\tau$-アファイン構造の立場から見れば，エスコート分布を考えるということが，$\tau$ の値を取り替えるということに相当している。つまり，測度空間の平行移動の仕方を取り替えることに相当しているのである。

第 9 章では，べき型の対数関数の性質として 2 種類の恒等式が成り立つことを示す。それに基づき，独立な確率変数 $X$ と $Y$ の同時確率分布とそれぞれの周辺分布を用いてエントロピーを表すと，優加法性になる場合と劣加法性になる場合が起こることを具体的に示す。第 10 章で重要な役割を演じることになる共形エントロピーとの相性がよいのは，劣加法性のほうである。また，相互情報量を具体的に計算することにより，ここでのエントロピーの非加法性が，確率変数の独立性と両立することも示される。非加法的エントロピーは，独立な確率変数に対して加法性を満たさないため，独立性の概念を拡張したものであると考える立場もあるが，本書では独立性の概念は不変なままであり，第 3 章で説明した十分統計量の概念を拡張したものとして，このことをとらえることにしている。

第 10 章では，劣加法性を持つエントロピーをスケール変換の自由度を用いることで加法的エントロピーに変換することを考える。ここで重要な役割を持つ

のはエントロピーそのものというよりは，共形エントロピーである。確率分布のスケール変換のもとで，この共形エントロピーがどのように振る舞うのかを調べることにより，また，どのようなスケール変換を行えば加法的（共形）エントロピーにできるのかを考えることにより，非加法性を加法性に変換できる。

第 11 章では，（共形）エントロピーの非加法性を加法性に変換するために利用されたスケール変換を表すパラメータを一つの座標とみなすことにより，ホログラフィー原理[23)]の一つの例として考える。このとき，計量は不定計量になっているが，これまで知られているような de Sitter 空間の計量とも Anti-de Sitter 空間の計量とも異なっていることが示される。また，ホログラフィー原理の例として非可換な演算子[†]に対してスケール変換のパラメータ（これを Trotter 座標と呼ぶ）を導入することで可換な演算子としてみなすことができるようにする鈴木-Trotter 変換を紹介する。また，ここでの $q$-積に対しても，べき型に拡張された指数関数に対して，鈴木-Trotter 変換と同様の方法により通常の指数関数の積とみなすことができることも示す。このように，1 次元だけ追加することで，状況が単純化し取扱いが容易になることを，ここではホログラフィー原理と呼んでいる。数学的には余次元 1 といってもよいようなものである。

第 12 章では，一般化平均である $\tau$-平均について考えることで，$\tau$-情報幾何学におけるパラメータ $\tau$ の値を決定する可能性について述べる。この $\tau$-平均は，Cooper による一般化平均[13)]と Hardy-Littlewood-Pólya により提案され Lin と Itô-Nara 等により発展させられた一般化平均の両方の性質を持っている[14)〜16)]。もし，得られたデータに対して（経験上）どのような平均を使えばよいのかがわかっている場合には，$\tau$-平均に含まれる二つのパラメータ $m$ と $\tau$ の値を決定することができる。これにより，$\tau$-情報幾何学に含まれるアファイン構造を決定するパラメータ $\tau$ の値が決まり，それに伴いデータの分布を特徴付ける確率分布族が決まることになる。

† ここでの演算子は，$e^X = \sum_{k=0}^{\infty} \frac{1}{k!} X^k$ のようにべき級数で定義される指数型の演算子を対象としている。

# 2 測度と確率

機械学習や深層学習の分野では，さまざまな場面で確率が登場する。その際，離散的な確率を扱う場合と連続的な確率を扱う場合が混在することも多い。このような場合には，確率変数が離散的な場合の確率を扱っているのか連続的な確率を扱っているのかに注意を払いつつ考えていく必要がある。ところが，測度論的確率論をほんの少し我慢して学んでみれば，離散の場合と連続の場合を統一的な枠組みで取り扱うことができるようになる。今後，人工知能などの分野も含め，ますます確率を取り扱う必要性は増していくことだろう。

そこで，この章では，測度論的確率論に関する基本的事項を説明する。しかし，すべてを取り上げるわけではなく，後の章で必要になる Radon-Nikodým の定理の紹介とその証明を与えることを目的として進めていくことにする。したがって，そのほかのことに関しては重要であっても省略しているので，興味のある方は他書を参照してもらいたい。

まず，さいころの例を用いて，後の章で必要になる用語を登場させた後，一般的な場合についてまとめる。ここで登場する用語は，試行，事象，根元事象，全事象，空事象，$\sigma$-加法族，可測関数，確率変数，可測空間，測度，測度空間，Borel $\sigma$-加法族，Lebesgue 測度などである。

## 2.1 可測空間と測度空間

ここでは，1 個のさいころを 1 回だけ投げるという単純な問題を少々厳密に考えることにより，後の章で必要になる用語と概念の導入を，冗長ではあるもの

# 索　　引

## —— 著 者 略 歴 ——

1986年　九州大学理学部物理学科卒業
1988年　九州大学大学院理学研究科修士課程修了（物理学専攻）
1991年　九州大学大学院理学研究科博士課程修了（物理学専攻），理学博士
1991年　電子技術総合研究所　研究員
1995年　電子技術総合研究所　主任研究官
1995年　大阪大学大学院助教授（連携大学院，～2000年）
2000年　埼玉大学助教授
2006年　福岡大学助教授
2007年　福岡大学准教授
2009年　福岡大学教授
　　　　現在に至る

# エントロピーの幾何学

Geometry of Entropy



2019 年 5 月 10 日　初版第 1 刷発行

検印省略

著　　者　田　中　　勝
発 行 者　株式会社　コ ロ ナ 社
　　　　　代 表 者　牛 来 真 也
印 刷 所　三 美 印 刷 株 式 会 社
製 本 所　有限会社　愛千製本所

112–0011　東京都文京区千石 4–46–10
発 行 所　株式会社　コ ロ ナ 社
CORONA PUBLISHING CO., LTD.
Tokyo Japan
振替 00140–8–14844・電話(03)3941–3131(代)
ホームページ　http://www.coronasha.co.jp

ISBN 978–4–339–02835–5　C3355　Printed in Japan　（齋藤）