統計数理 (1995)
第43巻 第2号 293-311

[原著論文]

分布の発展

――ルジャンドル変換と正準情報量規準――

統計数理研究所 松縄 規

(1994年11月 受付)

要旨

統計基礎モデルが与えられているものとして，その有様を記述するために導入される状態関数及びパラメータの変化によって，統計モデルがどう変化するかを，統計科学の見地からのLegendre変換とそれに係わるKullback–Leibler (K–L) 情報量の性質を主要な道具立てとして考究する．関連して，最小K–L情報量を基礎に，ある与えられた可測関数に関する正準分布が極く自然に定義され得ることを示す．また上記の変換を導入する際に必要となるK–L情報量の近似評価として，Kullbackの情報不等式および下半連続性を多変量の場合に拡張して検討する．更に，基礎確率分布のキュミュラント母関数の多変量Legendre変換が最大移動情報量および最小内部生成情報量と深く関わっていることを示す．この変換を軸に統計理論と応用分野との類推性，双対性を考察する．これらの結果を基に多変量の場合の正準情報量規準を考える．更にこの変換を逐次使用して基礎モデルを逐次改善・発展させるためのアルゴリズムを与える．それらの幾何学的解釈，線形計画問題における内点法との関連についても触れる．

1. はじめに

$\boldsymbol{A}$ を実確率空間 $(R^{n\times n}, \boldsymbol{B}^{n\times n}, P)$ 上で定義される $n\times n$ ランダム行列とする．以後の議論でこの空間又は確率分布 P を統計的基礎モデルと呼ぶ．今，定義域が $\boldsymbol{B}^{n\times n}$，その部分 σ–集合体の一つを値域とし平均的条件を満たすある可測関数が，P の拘束の下で，モデルのパラメータと関連して変化し，基礎モデルが新しい統計モデルへと変化する状況を考える．このことを本稿ではモデルの発展あるいは更新と呼ぶことにする．そのような変化をどう捉え，最適な統計モデルをいかに構築するかは興味ある重要課題である．この問題に従来の統計学の枠を越えた接近を試みることが本研究の目的である．

現代科学に於て統計的方法はデータ処理でそれなりに結果を出すため様々な分野で用いられている．しかし統計学の理論そのものは現在精力的な発展を見せている複雑性の科学をはじめとする諸科学で本質的に重要なものとして扱われているとは言い難いように見える．理由の一つは統計理論が諸科学からも類推の効く共通の用語で語ってこなかった為の他分野の無関心にあると思われる．本稿ではその点を考慮し他の応用分野，特に熱力学の状態関数の変化の理論と密接に関係させてある．この部分の考察は専門分野の違いと判断され組織的に考察されることがあまりなかった．これらの間の関係を明確にすることにより互いの分野の理論を系統的に理解するのに役立つ．そのため本稿では重要な統計的諸量とその熱力学的諸量との対応をつける必要から，数箇所で統計学の分野では聞き慣れない用語，例えば統計的比内部エネルギー，

内部生成情報などを導入する．これは統計学の背後に密接に関連する応用分野の具体的な理論と両者に共通な数理の存在を認識できると考えるからである．エントロピーや自由エネルギーといった用語が統計や確率論で使われるようになった今日，本稿で導入される用語に違和感を持たれないと思う．むしろ用語の語感から具体的次元を持つ量を連想し，モデルの構築の際の直感的理解を助ける利点もある．勿論，統計理論にのみ関心がある読者は，用語を気にせず記号のみに着目して読んで頂ければそれなりに理解が可能と思う．

さて冒頭の問題に取り組むために本稿では多変量の Kullback–Leibler 情報量をモデルの変化の評価量として導入し関連諸科学と整合する正準情報量規準を提案する．そのための重要な道具立てとして Legendre 変換を用いる．この変換は数学（微分方程式論，凸関数解析，確率論），解析力学，熱力学等で重要な変換であり，各分野の関連主要部分を相互理解するための強力な理論上の武器である．特に熱力学に於ける Legendre 変換は，K–L 情報量が負のエントロピーに相当することから，統計学における基礎モデルの更新の考察にあたり重要な類推的接近法を提供してくれることが期待される．ところでこれまでの統計学に於てもこの変換はそれと気づかれずに Kullback（1952, 1959）によって与えられた情報不等式の中に明確に現われ，それなりの利用がなされていた．

また Kullback and Leibler（1951）は数学的にはより一般的な不等式を与えていたがその証明中にいわゆる K–L 情報量間のピタゴラスの関係を得ていた．彼ら自身その重要性に気がつかなかったが，実はその中に現代の熱力学に於ける内部生成エントロピーに対応する情報量を得ていたと解釈でき，Prigogine and Defay（1954）による熱力学第二法則の新しい表現に係わる情報量による結果が潜在的にではあるが統計学の分野でも現われていたのは非常に興味深い．従来これらの事は指摘されたことがなかったと思われる．Kullback の結果は，不運なことにそれらの重要性が理解されないまま放置されて来たが，再評価が必要であろう．このような認識を伏線に上記課題を考察する．

K–L 情報量は色々な解釈が可能であるが，本稿では基本的にそれを基礎のモデル分布から発展分布への情報の移動と生成の両面を持つ汎関数という見方をして，分布間の離れ具合（あるいは距離の平方）という捉え方をしない．類似距離といった捉え方は問題によっては非常に有効であるが，K–L 情報量をエントロピーと関連づけてそれを徹底して生かした議論を展開しようとする時，本稿に於ける K–L 情報量の解釈は従来のそれに比べて有用さの点に於いて劣らない．

本稿の各章の概要は以下の通りである．次章で最小 K–L 情報量に基づく正準分布の導入を抽象測度空間上で示す．第3章では多変量の Kullback 情報量不等式と，等号条件としての情報量収支（or ピタゴラスの定理）の関係式が論じられている．それらの結果と Legendre 変換との関係，その幾何学的意味そして二つの本質的な情報量として移動情報量と内部生成情報量について触れている．第4章では以後の Legendre 変換の数学的議論で必要となる K–L 情報量の下半連続性が扱われている．第5章に本稿の主要結果である Legendre 変換とそれから導かれる正準情報量規準について論じてある．またそこでの結果を踏まえて，分布の逐次発展のアルゴリズムを第6章で扱う．

2. 最小 K–L 情報量に基づく正準分布

本章では K–L 情報量が統計理論に於てどのような意味を持っているかを示す一つの基本的性質を述べる．以下に記す結果は第5章で扱う Legendre 変換とそれに基づく正準情報量規準の抽象的一般化と見做し得るが，本稿ではむしろ正準分布の一般的把握と次章の統計基礎モデ

ルへの積率母関数導入の一つの根拠とすることを意図している．

$(\Omega,\ \mathscr{F},\ P)$ を基礎の確率空間とする．ここで Ω は抽象空間，$\mathscr{F}$ は Ω の部分集合からなる σ-集合体，P を基礎確率測度とする．$\rho(\omega)$ を可測空間 $(\Omega,\ \mathscr{F})$ 上で定義される非負の確率変数とし，次の関係を満たす確率測度 Q を考える：

$$dQ(\omega)=\rho(\omega)\,dP(\omega). \tag{2.1}$$

又，$\mathscr{B}$を $\rho(\omega)$ によって誘導される$\mathscr{F}$の部分 σ-集合体とする．この時上に対応して

$$dQ_{\mathscr{B}}(\omega)=E(\rho|\mathscr{B})\,dP_{\mathscr{B}}(\omega)$$

と表す．ここで $E(\rho|\mathscr{B})$ は$\mathscr{B}$が与えられた時の ρ の条件付き期待値，$P_{\mathscr{B}}$ は P の$\mathscr{B}$への制限，$Q_{\mathscr{B}}$ は

$$Q_{\mathscr{B}}(G)=\int_G E(\rho|\mathscr{B})\,dP_{\mathscr{B}}(\omega)=\int_G \rho(\omega)\,dP(\omega),\quad G\in\mathscr{B}$$

で与えられる．この時 Kullback（(1959)，p.20）を基に次のことが成立する：

命題 2.1. 基礎確率空間を $(\Omega,\ \mathscr{F},\ P)$ とし，$\rho(\omega)$ を (2.1) を満たす非負値を取る与えられた確率変数とする．また，$q(\omega)$ を関数型が未知の，確率測度 P に関する Radon–Nikodym 導関数を表す．この時$\mathscr{B}$が与えられた下での，$\rho(\omega)$ の P に関する条件付き密度関数は

$$I^*(\mathscr{F};q,\rho|\mathscr{B})=\int_\Omega q(\omega)\ln\left\{\frac{q(\omega)/E(q|\mathscr{B})}{\rho(\omega)}\right\}dP(\omega),\quad \omega\in\Omega \tag{2.2}$$

を最小にするものとして次のように与えられる：

$$q(\omega|\mathscr{B},\rho)=\rho(\omega)/E(\rho|\mathscr{B}). \tag{2.3}$$

逆に (2.3) が成立すれば $I^*(\mathscr{F};q,\rho\mid\mathscr{B})$ は最小となる．

証明.

$$\begin{aligned}I^*(\mathscr{F};q,\rho|\mathscr{B})&=\int_\Omega q(\omega)\ln\left\{\frac{q(\omega)/E(q|\mathscr{B})}{\rho(\omega)}\right\}dP(\omega)\\&=\int_\Omega \frac{q(\omega)}{\rho(\omega)}\ln\left[\frac{q(\omega)}{\rho(\omega)}\Big/\frac{E(q|\mathscr{B})}{E(\rho|\mathscr{B})}\right]dQ(\omega)-\int_\Omega q(\omega)\ln E(\rho|\mathscr{B})\,dP(\omega).\end{aligned}$$

第 1 積分は$\mathscr{B}$が与えられた時の条件付き修正情報量で非負値である．なぜなら

$$\int_\Omega \frac{q(\omega)}{\rho(\omega)}dQ(\omega)=\int_\Omega q(\omega)\,dP=1$$

および

$$\int_\Omega \frac{E(q|\mathscr{B})}{E(\rho|\mathscr{B})}dQ_{\mathscr{B}}(\omega)=\int_\Omega E(q|\mathscr{B})\,dP_{\mathscr{B}}(\omega)=\int_\Omega q(\omega)\,dP=1.$$

よって

$$I^*(\mathscr{F};q,\rho|\mathscr{B})\geq \underset{P}{E}[-\ln E(\rho|\mathscr{B})].$$

ここで等号は

$$\frac{q(\omega)}{E(q|\mathscr{B})} = \frac{\rho(\omega)}{E(\rho|\mathscr{B})} =: q(\omega|\mathscr{B}, \rho)$$

の時その時に限る. □

次の結果は上記命題の設定を修正して得られるが，この形での K–L 情報量の使用が便利なことがあるので独立して挙げる:

命題 2.2. $\eta(\omega)$ を測度空間 (Ω, $\mathscr{F}$, μ) で定義される与えられた可測関数とする. ここで Ω は抽象空間, $\mathscr{F}$ は Ω の部分集合からなる σ–集合体, μ は可測空間 (Ω, $\mathscr{F}$) 上の σ–有限測度を表す. 今, $A(\in\mathscr{F})$ を固定された集合とし，そこで $0<\eta(\omega)<\infty$ とする. また, $q_A(\omega)$ を A を台とする, 関数型が未知の, μ に関して絶対連続な任意の確率密度関数を表すものとする. この時 $q_A(\omega)$ の中で次の修正情報量

$$I^*(q_A;\eta|A) = \int_A q_A(\omega) \ln \frac{q_A(\omega)}{\eta(\omega)} d\mu$$

を最小にする分布として，与えられた A の下での $\eta(\omega)$ の μ に関する正準分布

$$q_A^*(\omega) = \frac{\eta(\omega)}{Z_A} \ [\mu], \qquad \left[Z_A = \int_A \eta(\omega)\, d\mu\right] \tag{2.4}$$

が得られる. 逆に $q_A(\omega) = q_A^*(\omega)$ であれば $I^*(q_A;\eta\mid A)$ は最小となる.

注 2.1. 上記最小値問題を Lagrange 乗数法を用いて形式的に解ける. 乗数を λ として汎関数微分を行なえば (cf. Boltzmann (1887))

$$\frac{\delta}{\delta q_A}\left[\int_A q_A(\omega) \ln \frac{q_A(\omega)}{\eta(\omega)} d\mu + \lambda\left(\int_A q_A(\omega)\, d\mu - 1\right)\right] = 0,$$

これより

$$\ln \frac{q_A^*(\omega)}{\eta(\omega)} + 1 + \lambda = 0,$$

即ち

$$q_A^*(\omega) = e^{-(\lambda+1)} \eta(\omega).$$

これが μ に関する確率密度関数となるために

$$e^{\lambda+1} = \int_A \eta(\omega)\, d\mu,$$

よって

$$q_A^*(\omega) = \eta(\omega) \Big/ \int_A \eta(\omega)\, d\mu(\omega).$$

この方法は，しかし，極値をとる為の必要条件のみに立脚しているという点で K–L 情報量に基づく方法に比べて限界がある.

3. 多変量分布に対する情報量不等式

R を実空間，$\boldsymbol{B}$ を R の Borel 集合族とし，$\boldsymbol{A}$ を $R^{n\times n}$ 上で定義されるランダム行列，P を $\boldsymbol{B}^{n\times n}$ 上の基礎の確率測度とする．$\boldsymbol{u}(\boldsymbol{A})$ を統計基礎モデル（$R^{n\times n}$, $\boldsymbol{B}^{n\times n}$, P）上で定義される $\boldsymbol{B}^{n\times n}$-可測な実数行列値関数とし，その平均に条件を付加し P を別の統計モデルに一種の入れ子構造で変化させることを考える．変化後のモデルを**発展モデル**と呼ぶことにする．換言すれば，発展モデルとはある条件の下 $\boldsymbol{u}(\boldsymbol{A})$ によって誘導される $\boldsymbol{B}^{n\times n}$ の部分 σ-集合体上への基礎モデルの更新された姿と言える．この変化を扱うため自然科学の中で最も基本的なものの一つとも言える熱力学の言葉を参考にして，統計学と熱力学の類推を容易にする用語を導入する．即ち，モデルの有様は統計的**状態量**と呼ぶ後述の特定の変数 $\boldsymbol{\beta}$, U, I 等で記述する．状態量という用語は様々な分野，例えばシステム工学や量子力学でそれぞれ独自の定義を持ち，重要な役割を演じているが，混乱することはないと思う．上記の変数と熱力学的状態量との類推は後で示される．状態量の変化を扱うために統計的**状態関数**を導入する．その一つとしてキュミュラント母関数を考え得る．状態関数を特定の状態量で偏微分することによりそれに関わる（共役）状態量と基本方程式が得られる（cf. (5.6)）．なお，Kapur（1989）は種々な事柄に加えて，統計学と熱力学の関連について本稿と違った側面から論じている．

ところで，可測関数 $\boldsymbol{u}(\boldsymbol{A})$ は基礎モデルを発展モデルへと変化させる，ランダム行列 1 個あたりが保有する，ある種の能力と見ることも出来る．従って $\boldsymbol{u}(\boldsymbol{A})$ を熱力学との類推で統計的**比内部エネルギー**（specific internal energy）と呼ぶのはある意味で自然である．以後の議論ではモデルのマクロな記述を対象としており，若干の条件の他は $\boldsymbol{u}(\boldsymbol{A})$ の関数型について特に制限をしない．これは熱力学に於て内部エネルギーが熱力学第一法則に従って定義される固有な概念であるが，それ以上に内容に立ち入らないのと同様である．考察対象の系のランダム要素間の相互作用を問題にする時は統計学と統計力学との類推を考える必要があるが本稿では扱わない．$\boldsymbol{u}(\boldsymbol{A})$ の基礎の確率測度 P に関するキュミュラント母関数が存在するものとして

$$\ln Z(\boldsymbol{\beta}) = \ln \int_{R^{n\times n}} \mathrm{etr}\,\{-\boldsymbol{\beta}^t \boldsymbol{u}(\boldsymbol{A})\}\, dP(\boldsymbol{A}), \tag{3.1}$$

と記す．ここで $\{\boldsymbol{\beta}^t : \boldsymbol{\beta}^t \in R^{n\times n},\ Z(\boldsymbol{\beta}) < \infty\}$ を考える．$\boldsymbol{\beta}$ は**正準パラメータ**と言われることもあるが，統計的**逆温度**（inverse temparature）と呼び得ることを後で示す．

さて，命題 2.1 を考慮して $\rho = \mathrm{etr}\{-\boldsymbol{\beta}^t \boldsymbol{u}(\boldsymbol{A})\}$（Boltzmann 因子）の P に関する正準分布を考え得る：任意の可測集合 $E \in \boldsymbol{B}^{n\times n}$ に対して

$$Q^*_{\boldsymbol{\beta}}(E) = \int_E \mathrm{etr}\,\{-\boldsymbol{\beta}^t \boldsymbol{u}(\boldsymbol{A})\}/Z(\boldsymbol{\beta})\, dP(\boldsymbol{A}). \tag{3.2}$$

即ち，$q^*_{\boldsymbol{\beta}} = dQ^*_{\boldsymbol{\beta}}(\boldsymbol{A})/dP(\boldsymbol{A})$ が任意の密度関数 q の中で，$\rho = \mathrm{etr}\{-\boldsymbol{\beta}^t \boldsymbol{u}(\boldsymbol{A})\}$ に対し情報量 $I^*(\boldsymbol{B}^{n\times n},\ q,\ \rho \mid \mathscr{B}[\boldsymbol{u}(\boldsymbol{A})],\ R^{n\times n})$ を最小にする分布である．ここに $\mathscr{B}[\boldsymbol{u}(\boldsymbol{A})]$ は $\boldsymbol{u}(\boldsymbol{A})$ が誘導する $\boldsymbol{B}^{n\times n}$ の部分 Borel σ-集合体を表す．この時 $Z(\boldsymbol{\beta}) = E(\rho \mid \mathscr{B}[\boldsymbol{u}(\boldsymbol{A})],\ R^{n\times n})$.

更に統計理論に於いて基本的な次の条件下で問題を考える：$\boldsymbol{u}(\boldsymbol{A})$ が $Q^*_{\boldsymbol{\beta}}$-可積分とし，

$$\boldsymbol{U}(\boldsymbol{\beta}) := \int_{R^{n\times n}} \boldsymbol{u}(\boldsymbol{A})\, dQ^*_{\boldsymbol{\beta}}(\boldsymbol{A}) \tag{3.3}$$

と表す．これは以後の巨視的記述に必要な量で統計的**内部エネルギー**（internal energy）と呼ぶ．

さて，Q を $\boldsymbol{B}^{n\times n}$ 上の任意の確率分布とし，K–L 情報量

$$(3.4)\qquad I(Q;P)=\int_{R^{n\times n}}[\ln\{dQ/dP\}]\,dQ \quad \text{if} \quad Q\ll P;$$
$$=\infty \quad \text{otherwise,}$$

を考える．特に $Q=Q_\beta^*$ の場合（3.1)–(3.3）により

$$(3.5)\qquad I(Q_\beta^*;P)=\int_{R^{n\times n}}[\ln\{dQ_\beta^*/dP\}]\,dQ_\beta^*=\int_{R^{n\times n}}\mathrm{tr}\{-\boldsymbol{\beta}^t\boldsymbol{u}(\boldsymbol{A})-\ln Z(\boldsymbol{\beta})\}dQ_\beta^*$$
$$=\mathrm{tr}\{-\boldsymbol{\beta}^t\boldsymbol{U}(\boldsymbol{\beta})\}-\ln Z(\boldsymbol{\beta})<\infty.$$

上式は実は情報量とキュムラント母関数が互いにLegendre双対の関係にあることを示している．このことのより詳しい議論は第5章で行なう．

我々はKullback（1952）の情報不等式を次のように多変量分布の場合へ拡張できる：

命題 3.1. Q は $\boldsymbol{B}^{n\times n}$ 上の確率分布で $\boldsymbol{u}(\boldsymbol{A})$ が Q–可積分かつ

$$(3.6)\qquad \int_{R^{n\times n}}\boldsymbol{u}(\boldsymbol{A})\,dQ(\boldsymbol{A})=\boldsymbol{U}$$

が満たされる時，不等式

$$(3.7)\qquad I(Q;P)\geq I(Q_\beta^*;P)=-\mathrm{tr}\{\boldsymbol{\beta}^t\boldsymbol{U}(\boldsymbol{\beta})\}-\ln Z(\boldsymbol{\beta})$$

が成立する．等号は $\boldsymbol{B}^{n\times n}$ 上で $Q=Q_\beta^*$ の時その時に限る．

証明. $I(Q;P)<\infty$ を仮定する．従って $Q\ll P$.

$$h(\boldsymbol{A})=dQ(\boldsymbol{A})/dP(\boldsymbol{A}),$$

と置けば

$$I(Q;P)=\int_{R^{n\times n}}h(\boldsymbol{A})\ln h(\boldsymbol{A})\,dP(\boldsymbol{A})$$

および

$$I(Q_\beta^*;P)=\int_{R^{n\times n}}h^*(\boldsymbol{A})\ln h^*(\boldsymbol{A})\,dP(\boldsymbol{A}),$$

ここに

$$h^*(\boldsymbol{A})=dQ_\beta^*(\boldsymbol{A})/dP(\boldsymbol{A})=\mathrm{etr}\{-\boldsymbol{\beta}^t\boldsymbol{u}(\boldsymbol{A})\}/Z(\boldsymbol{\beta}).$$

従って

$$\int_{R^{n\times n}}h(\boldsymbol{A})\ln h^*(\boldsymbol{A})\,dP(\boldsymbol{A})=\int_{R^{n\times n}}\ln h^*(\boldsymbol{A})\,dQ(\boldsymbol{A})=-\mathrm{tr}\{\boldsymbol{\beta}^t\boldsymbol{U}\}-\ln Z(\boldsymbol{\beta})$$
$$=\int_{R^{n\times n}}\ln h^*(\boldsymbol{A})\,dQ_\beta^*(\boldsymbol{A})=\int_{R^{n\times n}}h^*(\boldsymbol{A})\ln h^*(\boldsymbol{A})\,dP(\boldsymbol{A})=I(Q_\beta^*;P).$$

$t>0$ に対し $-\ln t\geq 1-t$（等号：$t=1$）が成立することおよび

$$\int_{R^{n\times n}}h(\boldsymbol{A})\,dP(\boldsymbol{A})=\int_{R^{n\times n}}h^*(\boldsymbol{A})\,dP(\boldsymbol{A})=1$$

に注意すれば

$$(3.8)\quad I(Q;P)-I(Q_\beta^*;P)$$
$$=\int_{R^{n\times n}} h(\boldsymbol{A})\ln\,[h(\boldsymbol{A})/h^*(\boldsymbol{A})]\,dP(\boldsymbol{A})\geq\int_{R^{n\times n}} h(\boldsymbol{A})\,[1-h^*(\boldsymbol{A})/h(\boldsymbol{A})]\,dP(\boldsymbol{A})=0,$$

ここに等号は $\boldsymbol{B}^{n\times n}$ 上で $Q=Q_\beta^*$ の時その時に限る．□

次の結果は本質的に Kullback and Leibler（1951）が彼らの定理 4.1 の証明の中で到達している．その後何人かの研究者により再発見されている（cf. 注 3.2)．ここでは多変量分布に対する表現を与える．情報量を疑似距離と見る立場では，次の等式はピタゴラスの定理と見做せる．本稿では 1 章で述べた情報の移動と生成という立場から，熱力学に於けるエントロピー収支（cf. Hiraoka and Tanaka（1994)）に倣って，情報量収支と言うことにする：

命題 3.2. 定理 3.1 の条件下で，情報量の収支関係が成立する：

$$(3.9)\qquad I(Q;P)=I(Q;Q_\beta^*)+I(Q_\beta^*;P).$$

証明.

$$(3.10)\qquad \int_{R^{n\times n}} h(\boldsymbol{A})\ln\frac{h(\boldsymbol{A})}{h^*(\boldsymbol{A})}dP(\boldsymbol{A})=\int_{R^{n\times n}}\frac{dQ}{dP}\ln\frac{dQ/dP}{dQ_\beta^*/dP}dP(\boldsymbol{A})$$
$$=\int_{R^{n\times n}}\ln\frac{dQ}{dQ_\beta^*}dQ=I(Q;Q_\beta^*)\geq 0.$$

(3.8)，(3.10) から直ちに所要の (3.9) を得る．□

注 3.1. 式 (3.9) には次の意味での直交性が含まれる．

$$(3.11)\qquad \int_{R^{n\times n}}\ln\frac{dQ_\beta^*}{dP}d(Q-Q_\beta^*)(\boldsymbol{A})=0.$$

何故ならば

$$I(Q;P)=\int_{R^{n\times n}}\ln\frac{dQ}{dP}dQ(\boldsymbol{A})=\int_{R^{n\times n}}\left[\ln\frac{dQ}{dQ_\beta^*}+\ln\frac{dQ_\beta^*}{dP}\right]dQ$$

一方，関係 (3.9) により

$$I(Q;P)=\int_{R^{n\times n}}\ln\frac{dQ}{dQ_\beta^*}dQ+\int_{R^{n\times n}}\ln\frac{dQ_\beta^*}{dP}dQ_\beta^*$$

従って

$$\int_{R^{n\times n}}\ln\frac{dQ_\beta^*}{dP}dQ=\int_{R^{n\times n}}\ln\frac{dQ_\beta^*}{dP}dQ_\beta^*.$$

式 (3.11) の意味を幾何学的直感から補足的に説明すると次のようになる．可測空間（$R^{n\times n}$, $\boldsymbol{B}^{n\times n}$）上で定義されるランダム行列 $\boldsymbol{A}$ の確率分布全体を$\mathbb{F}_0$とし，条件 (3.6) を満たす確率分布部分族$\mathbb{F}_1$が存在するとする．明らかに $P\in\mathbb{F}_0$，$\{Q, Q_\beta^*\}\subset\mathbb{F}_1$である．これらのことを考慮して我々は次の様な一つの情報空間$\mathscr{I}_0$の幾何学的イメージを描ける．$\mathbb{F}_0$に於ける確率分布を$\mathscr{I}_0$の点とし，P に拘束され$\mathbb{F}_1$での平均条件 (3.6) に対応する $d(Q-Q_\beta^*)(\boldsymbol{A})$ で定まる方向場を持つ被拘束超平面を$\mathscr{U}_1$とする．点 P を通り $\ln[dQ_\beta^*/dP]$ で方向づけられる超平面に含まれる情報測地線$\mathscr{g}_1$を考えれば$\mathscr{g}_1$と$\mathscr{U}_1$は (3.11) の意味で直交する．この時$\mathscr{g}_1$の$\mathscr{U}_1$への足が Q_β^* を与える．また，P から Q への情報測地線へ $I(Q;P)$ を，P から Q_β^* への情報測地線$\mathscr{g}_1$に $I(Q_\beta^*;P)$，

Q_β^* から Q への情報測地線へ $I(Q;Q_\beta^*)$ を対応させると，これらの間に命題3.2に見るように情報空間に於ける情報収支の関係が成立する．しかしK–L情報量は一般に対称性及び三角公理を満たさないからここでの情報空間は距離空間とはならない．なお，Csiszár (1975) は (3.9) に関して，情報射影の概念を導入した後 Hilbert 空間における理論との類推から導いている．また Ghurye (1968) は条件付き情報の概念を導入してこれを与えている．離散型分布に関して Kapur and Kesavan (1992) も (3.9) 式を与えている．

注 3.2. 式 (3.10) から $I(Q;Q_\beta^*)$ は基礎モデル P とは陽には相互依存関係を持たない形で与えられている．一方，命題3.1の不等式 (3.7) と命題3.2の不等式 (3.9) を比較すると，一般に情報量は保存されないが $\boldsymbol{B}^{n\times n}$ 上で $Q=Q_\beta^*$ の時に限って情報量の釣合式が成立することが分かった．ところで分布 $\{Q,\ Q_\beta^*\}$ の外部に基礎モデル P 以外の分布を考えていないから，不等式 (3.9) を等式化している $I(Q;Q_\beta^*)$ は**内部生成情報量** (internal produced information) と言ってよい．これは Prigogine and Defay (1954) による熱力学に於けるエントロピー生成から類推される量である．また $I(Q;Q_\beta^*)\geq 0$ は $\boldsymbol{u}(\boldsymbol{A})$ が誘導する $\boldsymbol{B}^{n\times n}$ の部分 Borel σ–集合体 $\mathscr{B}[\boldsymbol{u}(\boldsymbol{A})]$ に関する Q の条件付き Q_β^*–情報量 (conditional Q_β^*–information of Q in $\boldsymbol{B}^{n\times n}$ relative to $\mathscr{B}[\boldsymbol{u}(\boldsymbol{A})]$) に対応しており，$I(Q;Q_\beta^*)=0$ の時 $\mathscr{B}[\boldsymbol{u}(\boldsymbol{A})]$ は $\boldsymbol{B}^{n\times n}$ の十分部分 Borel σ–集合体 (Ghurye (1968))，従って $\boldsymbol{u}(\boldsymbol{A})$ は $\boldsymbol{B}^{n\times n}$ 上の確率分布 P に関して十分統計量となる (Halmos and Savage (1949))．

注 3.3. 式 (3.10) の $I(Q_\beta^*;P)$ は基礎モデル P からモデル Q_β^* を構築する際に P から Q_β^* への**移動情報量** (transported information) と解釈される．これは熱力学に於ける移動 (or 輸送) エントロピーからの類推である (cf. Prigogine and Defay (1954))．命題3.1からこの情報量がモデル更新に於ける実質的部分であることが分かる．またこれが実は Legendre 変換になっていることを第5章で明らかにする．即ち不等式 (3.7) の下限が式 (5.7) に現われてくる．また $I(Q_\beta^*;P)$ は部分 Borel σ–集合体 $\mathscr{B}[\boldsymbol{u}(\boldsymbol{A})]$ に含まれる情報量を表している (cf. Kallianpur (1960), Ghurye (1968))．

注 3.4. 命題3.2の諸分布の間に次の結果が成立する：もし

$$\int_{R^{n\times n}} \ln\frac{dQ_\beta^*}{dQ}\,d(P-Q_\beta^*)(\boldsymbol{A})=0 \tag{3.12}$$

が満たされるならば，$I(Q;P)$ の双対情報 $I(P;Q)$ に対し次の関係式が成り立つ：

$$I(P;Q)=I(P;Q_\beta^*)+I(Q_\beta^*;Q). \tag{3.13}$$

4. K–L 情報量の下半連続性

本章ではK–L情報量の代表的極限性質，汎関数の下半連続性を扱う．この性質は次章の Legendre 変換に於いて，また一般に分布の大偏差確率の考察，統計的エントロピー関数の導入時等に本質的に重要な性質である．以下に次章との関連で Kullback ((1959), p.70) とは異なる設定での結果を与える．

$\mathscr{P}(R^{n\times n},\ \boldsymbol{B}^{n\times n},\ \mu)$ を $(R^{n\times n},\ \boldsymbol{B}^{n\times n})$ 上の σ–有限測度 μ に関し絶対連続な全ての確率分布の族とする．$\{P_\nu\}\,(\nu=1,2,\ldots)$，$\{Q_\nu\}\,(\nu=1,2,\ldots)$ を $\mathscr{P}(R^{n\times n},\ \boldsymbol{B}^{n\times n},\ \mu)$ に属する二つの確率分

布列，P および Q もこの族に入る確率分布とし次の極限過程を満たすものと仮定する：

(4.1) $$\lim_{\nu\to\infty} P_\nu(E)=P(E), \qquad \lim_{\nu\to\infty} Q_\nu(E)=Q(E), \qquad \text{for } E\in \boldsymbol{B}^{n\times n} \text{ and } Q\ll P.$$

命題 4.1. 二つの確率分布の列 $\{P_\nu\}$ $(\nu=1,2,\ldots)$，$\{Q_\nu\}$ $(\nu=1,2,\ldots)$ が条件 (4.1) を満たすならば，不等式

(4.2) $$\liminf_{\nu\to\infty} I(Q_\nu;P_\nu)\geq I(Q;P)$$

が成立する．

証明. P_ν，$Q_\nu\in\mathscr{P}(R^{n\times n},\ \boldsymbol{B}^{n\times n},\ \mu)$ だから任意の $E\in\boldsymbol{B}^{n\times n}$ に対して

$$P_\nu(E)=\int_E p_\nu(\boldsymbol{A})d\mu(\boldsymbol{A}) \quad \text{and} \quad Q_\nu(E)=\int_E q_\nu(\boldsymbol{A})d\mu(\boldsymbol{A})$$

と表現できる．ここに p_ν 及び q_ν は夫々 μ に関する P_ν，Q_ν の Radon-Nikodym 導関数を表す．よって，K–L 情報量を次のように与え得る：

$$I(Q_\nu;P_\nu)=\int_{\boldsymbol{B}^{n\times n}} q_\nu(\boldsymbol{A})\ln\frac{q_\nu(\boldsymbol{A})}{p_\nu(\boldsymbol{A})}d\mu(\boldsymbol{A}).$$

ここで $q_\nu/p_\nu=h_\nu$ と置くと $dQ_\nu=q_\nu d\mu=h_\nu p_\nu d\mu=h_\nu dP_\nu$．よって

$$I(Q_\nu;P_\nu)=\int_{\boldsymbol{B}^{n\times n}} h_\nu(\boldsymbol{A})\ln h_\nu(\boldsymbol{A})\,dP_\nu(\boldsymbol{A})=\int_{\boldsymbol{B}^{n\times n}} dQ_\nu\ln\frac{dQ_\nu}{dP_\nu}.$$

さて K–L 情報量の定義により

$$I(Q_\nu;P_\nu)=\sup_{\{E_i\}}\sum_i\int_{E_i} h_\nu(\boldsymbol{A})\ln h_\nu(\boldsymbol{A})\,dP_\nu(\boldsymbol{A}).$$

ここに $\{E_i\}$ は可能な全ての $R^{n\times n}$ の $\boldsymbol{B}^{n\times n}$–可測な有限分割を表す．Jensen の不等式を用いて

$$\geq\sup_{\{E_i\}}\sum_i\left[\left(\int_{E_i}h_\nu dP_\nu\right)\ln\frac{\int_{E_i}h_\nu dP_\nu}{\int_{E_i}dP_\nu}\right]=\sup_{\{E_i\}}\sum_i\left[\left(\int_{E_i}dQ_\nu\right)\ln\frac{\int_{E_i}dQ_\nu}{\int_{E_i}dP_\nu}\right].$$

即ち，

$$I(Q_\nu;P_\nu)\geq\sup_{\{E_i\}}\sum_i Q_\nu(E_i)\ln\frac{Q_\nu(E_i)}{P_\nu(E_i)}.$$

右辺は確率分布の連続関数であることに注意して条件 (4.1) を使えば

(4.3) $$\liminf_{\nu\to\infty} I(Q_\nu;P_\nu)\geq\sup_{\{E_i\}}\sum_i Q(E_i)\ln\frac{Q(E_i)}{P(E_i)}.$$

一方，K–L 情報量 $I(Q;P)$ は

(4.4) $$\sup_{E_i\in\boldsymbol{B}^{n\times n}}\sum_i Q(E_i)\ln\frac{Q(E_i)}{P(E_i)}=\int q(\boldsymbol{A})\ln\frac{q(\boldsymbol{A})}{p(\boldsymbol{A})}d\mu(\boldsymbol{A})$$

と積分表示が可能だから (cf. Kallianpur(1960)，Ghurye(1968))，結局 (4.3)，(4.4) により所要の結果

$$\liminf_{\nu\to\infty}\int q_\nu(\boldsymbol{A})\ln\frac{q_\nu(\boldsymbol{A})}{p_\nu(\boldsymbol{A})}d\mu(\boldsymbol{A})\geq\int q(\boldsymbol{A})\ln\frac{q(\boldsymbol{A})}{p(\boldsymbol{A})}d\mu(\boldsymbol{A})$$

を得る. □

注 4.1. Dobrushin (1960) は本定理の原形を条件 (4.1) よりも強い変動距離の意味での収束条件の下で証明している. なお, この定理は積分論に於ける Fatou の補題の一般化とも見做せる.

5. Legendre 変換と正準情報量規準

本章では第3章と同様基礎の確率空間を $(R^{n\times n},\ \boldsymbol{B}^{n\times n},\ P)$ とする. 基礎モデル P の有様が統計的状態変数 $\boldsymbol{\beta}$ や $\boldsymbol{U}$ の変動によりどのように変化するかということは前述したように非常に興味のある問題である. これに対し正準情報量極小化を原理として考え得ることを示す. この為の重要な数理的道具として Legendre 変換 (Rockafellar (1970)) が利用できる. この変換は Legendre 以前, 少なくとも A.C. Clairaut (1713–1765) の微分方程式 $y=px+f(p)$, $p=dy/dx$ の解及び空間曲線の研究に現われていた (cf. Arnold (1980)). しかし, よく知られている Legendre 変換は解析力学に於ける Lagrange-Hamilton 双対と熱力学に現われる重要な状態関数間の変換であろう. 特に後者に関連して熱力学的エントロピーと情報量の関係が明らかになるにつれ, 大偏差確率の理論の中でもこの変換の重要さが指摘されている (cf. Brown (1986), Deuscel and Strook (1989), Barndorff-Nielsen and Cox (1989)). 以上のことから Legendre 変換を用いて様々な切り口から上の問題を扱える. 一般に, 他分野の基本的問題を統計的観点から, 或は逆に統計の基本問題を関連科学の観点から系統的に理解できる事も多々あると思われる. 以下主として統計理論と熱力学理論の類推を念頭に考察する. この点を強調する為第3章でのキュミュラント母関数を統計的**特性関数** (statistical characteristic function) と呼び

$$\Psi(\boldsymbol{\beta})=\ln Z(\boldsymbol{\beta})=\ln\int_{R^{n\times n}}\mathrm{etr}\{-\boldsymbol{\beta}^t\boldsymbol{u}(\boldsymbol{A})\}\,dP(\boldsymbol{A})\ (<\infty)\tag{5.1}$$

と表すことにする. ここで $-\Psi(\boldsymbol{\beta})$ の Legendre 変換を次のように定義する:

$$S(\boldsymbol{u}):=\inf_{\boldsymbol{\beta}\in\Gamma}\{\mathrm{tr}(\boldsymbol{\beta}^t\boldsymbol{u})-(-\Psi(\boldsymbol{\beta}))\}=\inf_{\boldsymbol{\beta}\in\Gamma}\{\Psi(\boldsymbol{\beta})+\mathrm{tr}(\boldsymbol{\beta}^t\boldsymbol{u})\},\tag{5.2}$$

ここに Γ は $Z(\boldsymbol{\beta})<\infty$ となる $\boldsymbol{\beta}$ の集合の内点を表す. Γ は空集合ではないものとする. また Γ に属さない Γ の境界点では $\Psi(\boldsymbol{\beta})=\infty$ とする.

上式の変換は M. Planck の熱力学的特性関数 (thermodynamical characteristic function) $\Psi_t=S_t-U_t/T$ (cf. Kubo (1952)) に対応している. 但し, 熱力学系は閉鎖しており, S_t は熱力学的エントロピー, U_t は熱力学的平均内部エネルギー, T は絶対温度を表す.

上のような設定の下で統計的特性関数の $\boldsymbol{\beta}$ に関する変化を考えよう:

$$\begin{aligned}\frac{\partial\Psi(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}}&=\left(\frac{\partial}{\partial\beta_{ij}}\right)\Psi(\boldsymbol{\beta})=-\frac{\int_{R^{n\times n}}\boldsymbol{u}(\boldsymbol{A})\,\mathrm{etr}\{-\boldsymbol{\beta}^t\boldsymbol{u}(\boldsymbol{A})\}\,dP(\boldsymbol{A})}{\int_{R^{n\times n}}\mathrm{etr}\{-\boldsymbol{\beta}^t\boldsymbol{u}(\boldsymbol{A})\}\,dP(\boldsymbol{A})}\\&=:-\int_{R^{n\times n}}\boldsymbol{u}(\boldsymbol{A})\,dQ^*_{\boldsymbol{\beta}}(\boldsymbol{A})=-\boldsymbol{U}(\boldsymbol{\beta}),\end{aligned}\tag{5.3}$$

ここで

(5.4) $$dQ^*_{\beta}(\boldsymbol{A}) = \mathrm{etr}\{-\boldsymbol{\beta}^t \boldsymbol{u}(\boldsymbol{A})\}/Z(\boldsymbol{\beta})\,dP(\boldsymbol{A}),\quad \boldsymbol{A}\in R^{n\times n},$$

と置いた．これはいわゆる正準分布であり，命題2.1により一つの合理的な選択である．行列 $\boldsymbol{U}(\boldsymbol{\beta})=(U_{ij})$ は我々の考察対象の状態関数で関数 $\boldsymbol{u}(\boldsymbol{A})$ の最適更新分布 Q^*_{β} による平均である．$\Psi(\boldsymbol{\beta})$ を更に $\boldsymbol{\beta}$ で微分して，以下の議論は行列が非退化の場合のみを扱うものとして，$\boldsymbol{u}$ の分散共分散行列の性質を考慮すれば

(5.5) $$\frac{\partial}{\partial\boldsymbol{\beta}}\left(\frac{\partial\Psi(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}}\right)^t = -\frac{\partial \boldsymbol{U}^t(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}} = \left(\frac{\partial}{\partial\beta_{ij}}\right)(-U_{ji}) \gg 0,\ \text{i.e. positive definite.}$$

従って $\Psi(\boldsymbol{\beta})$ は，二次導関数が上記の意味で正定値であるスカラー値多変量狭義凸関数であり，写像 $\boldsymbol{\beta}\to\boldsymbol{U}$ は 1：1 かつ無限回微分可能な行列値多変量関数である．従ってこの写像は Γ から関数 $S(\boldsymbol{u})$ の定義域 G_S 上への 1：1 対応でその逆写像はまた 1：1 かつ無限回微分可能な行列値多変量関数である．即ち $\Psi(\boldsymbol{\beta})$ の $\boldsymbol{\beta}$ に関する導関数は $R^{n\times n}$ のそれ自身への可微分同相写像である．よって Legendre 変換（5.2）は狭義凸関数でその下限は

(5.6) $$\frac{\partial}{\partial\boldsymbol{\beta}}(-\Psi(\boldsymbol{\beta})) = \boldsymbol{U}(\boldsymbol{\beta})$$

を満たす唯一の点 $\boldsymbol{\beta}=\boldsymbol{\beta}(\boldsymbol{U})$ で達成される．この時（5.2）から，統計的**エントロピー関数**が

(5.7) $$S(\boldsymbol{U}) := \Psi(\boldsymbol{\beta}) + \mathrm{tr}(\boldsymbol{\beta}^t\boldsymbol{U})$$

で定義される．

さて，ここで統計的エントロピー関数 $S(\boldsymbol{U})$ の別の表現を考えよう．（5.4）から

$$-\ln\frac{dQ^*_{\beta}(\boldsymbol{A})}{dP(\boldsymbol{A})} = \mathrm{tr}\{\boldsymbol{\beta}^t\boldsymbol{u}(\boldsymbol{A})\} + \ln Z(\boldsymbol{\beta}).$$

上式の両辺の $dQ^*_{\beta}(\boldsymbol{A})$ に関する平均を考えると

(5.8) $$-\int_{R^{n\times n}}\ln\frac{dQ^*_{\beta}(\boldsymbol{A})}{dP(\boldsymbol{A})}dQ^*_{\beta}(\boldsymbol{A}) = \mathrm{tr}\left\{\boldsymbol{\beta}^t\int_{R^{n\times n}}\boldsymbol{u}(\boldsymbol{A})\,dQ^*_{\beta}(\boldsymbol{A})\right\} + \ln Z(\boldsymbol{\beta})$$
$$= \mathrm{tr}\{\boldsymbol{\beta}^t\boldsymbol{U}(\boldsymbol{\beta})\} + \Psi(\boldsymbol{\beta}),$$

となる．（5.7）と比較して直ちに

(5.9) $$S(\boldsymbol{U}) = -\int_{R^{n\times n}}\ln\frac{dQ^*_{\beta}(\boldsymbol{A})}{dP(\boldsymbol{A})}dQ^*_{\beta}(\boldsymbol{A})\ \ (=\text{minus K–L information } I(Q^*_{\beta};P)).$$

即ち，K–L 情報量 $I(Q^*_{\beta};P)$ は $S(\boldsymbol{U})$ のネゲントロピー（negentropy, negative-entropy）であることが確認された．これは一変量の場合に Brillouin（1956）が直感的に説明したことの統計数理面からの証明である．この結果と命題4.1による $I(Q^*_{\beta};P)$ の下半連続性とからエントロピー関数 $S(\boldsymbol{U})$ の上半連続性が直ちに得られる．

従って，我々は統計的特性関数を次の（5.2）と共役な Legendre 変換

(5.10) $$\Psi(\boldsymbol{\beta}) = \sup_{\boldsymbol{u}\in G_S}\{S(\boldsymbol{u}) - \mathrm{tr}(\boldsymbol{\beta}^t\boldsymbol{u})\}$$

によって表現できる（Rockafellar（1970））．これは熱力学に於ける Planck の平衡条件に相当

する (cf. Kubo (1952)). 命題3.1の不等式 (3.7) の下限はこの平衡条件に対応する，情報量を用いた統計的表現と見なせる.

本章のここまでの議論と命題3.1から基礎モデルPの最適発展モデルQ_β^*を求めるための正準情報量規準を得る:

定理 5.1. 確率分布Pを $(R^{n\times n}, \boldsymbol{B}^{n\times n})$ 上で定義される統計基礎モデル，$\boldsymbol{u}(\boldsymbol{A})$ を実数行列値を取る$\boldsymbol{B}^{n\times n}$–可測関数，$\boldsymbol{U}$を$\boldsymbol{u}$の本質的値域の凸包の内点とする. この時，$P$の最適発展モデルは次の正準情報量規準を満たす: 命題 (a)

$$I(Q_\beta^*;P)=\inf_Q\left\{I(Q;P):\int_{R^{n\times n}}\boldsymbol{u}(\boldsymbol{A})\,dQ(\boldsymbol{A})=\boldsymbol{U}<\infty\right\}. \tag{5.11}$$

上で下限は$Q=Q_\beta^*$の時またその時に限り，$\boldsymbol{\beta}$は次式を満たすように選ばれる:

$$\int_{R^{n\times n}}\boldsymbol{u}(\boldsymbol{A})\,dQ_\beta^*(\boldsymbol{A})=\boldsymbol{U}\ ; \tag{5.12}$$

またはPの最適発展モデルは同等な命題 (b)

$$I(Q_\beta^*;P)=-\inf_{\beta\in\Gamma}\{\Psi(\boldsymbol{\beta})+\mathrm{tr}(\boldsymbol{\beta}^t\boldsymbol{u})\} \tag{5.13}$$

を満たす. ここで$\boldsymbol{\beta}$は次式が成立するように選ばれる:

$$\frac{\partial}{\partial\boldsymbol{\beta}}(-\Psi(\boldsymbol{\beta}))=\boldsymbol{U}(\boldsymbol{\beta}). \tag{5.14}$$

また，(a)，(b) のどちらの場合も基礎モデルPの最適発展モデルQ_β^*の Radon-Nikodym の導関数は次式で与えられる:

$$dQ_\beta^*(\boldsymbol{A})=\frac{\mathrm{etr}\{-\boldsymbol{\beta}^t\boldsymbol{u}(\boldsymbol{A})\}}{Z(\boldsymbol{\beta})}dP(\boldsymbol{A}),\quad \boldsymbol{A}\in R^{n\times n}. \tag{5.15}$$

ここで$Z(\boldsymbol{\beta})$は (5.1) で定義される$\boldsymbol{u}$のPに関する積率母関数を表す.

注 5.1. 統計的エントロピー関数$S(\boldsymbol{U})$の凹性又は同等な情報量$I(Q_\beta^*;P)$の凸性は次の様にしても従う: $S(\boldsymbol{U})=\Psi(\boldsymbol{\beta})+\mathrm{tr}(\boldsymbol{\beta}^t\boldsymbol{U})=\ln Z(\boldsymbol{\beta})+\mathrm{tr}(\boldsymbol{\beta}^t\boldsymbol{U})$ with $\boldsymbol{\beta}=\boldsymbol{\beta}(\boldsymbol{U})$ に注意し，

$$\begin{aligned}\frac{\partial S(\boldsymbol{U})}{\partial\boldsymbol{U}}&=:\left(\frac{\partial}{\partial U_{ij}}\right)S(\boldsymbol{U})=\left(\frac{\partial\ln Z(\boldsymbol{\beta})}{\partial\beta_{ij}}\right)\left(\frac{\partial\beta_{ij}}{\partial U_{ji}}\right)+\boldsymbol{\beta}+\boldsymbol{U}\left(\frac{\partial\beta_{ij}}{\partial U_{ji}}\right)\\&=-\boldsymbol{U}\frac{\partial\boldsymbol{\beta}}{\partial\boldsymbol{U}^t}+\boldsymbol{\beta}+\boldsymbol{U}\frac{\partial\boldsymbol{\beta}}{\partial\boldsymbol{U}^t}=\boldsymbol{\beta}.\end{aligned} \tag{5.16}$$

この事と熱力学との類推から，$\boldsymbol{\beta}$は統計的逆温度と呼び得る. (5.16) を微分し (5.5) を考慮すれば

$$\frac{\partial}{\partial\boldsymbol{U}}\left(\frac{\partial S(\boldsymbol{U})}{\partial\boldsymbol{U}}\right)^t=\frac{\partial\boldsymbol{\beta}^t}{\partial\boldsymbol{U}}=\left(\frac{\partial\boldsymbol{U}}{\partial\boldsymbol{\beta}^t}\right)^{-1}\ll 0,\ \text{i.e. negative definite.} \tag{5.17}$$

即ち，統計的エントロピー関数は狭義凹関数である.

注 5.2. (5.15) は Bayes の定理の一つの表現でもある. 即ち，dPは事前分布 (＝統計基礎モデル) の微分，$\mathrm{etr}\{-\boldsymbol{\beta}^t\boldsymbol{u}\}/Z(\boldsymbol{\beta})$ は$\boldsymbol{u}$によって誘導される部分σ–集合体が与えられた下での，

$\rho = \mathrm{etr}\{-\boldsymbol{\beta}^t\boldsymbol{u}\}$ の P に関する条件付確率密度であり，従って dQ_{β}^* は対応する事後分布の微分に対応する．

注 5.3. 本定理の規準を適用するにあたって基礎統計モデルの P を知ることが本質的である．Kullback はこの点で何も明らかにしなかった．彼の諸結果がなかなか利用されなかった大きな理由の一つもここにある．これに対処するために例えば，統計的不確定性関係に基づく統計基礎方程式の利用が考え得る（cf. Matsunawa（1994））．

注 5.4. 統計的平均エネルギー U を最適に調節するためには，例えばそれを $\boldsymbol{A}$ の一般化算術平均:

$$\tilde{\boldsymbol{U}} = \phi_{\boldsymbol{U}}^{-1}\left(n^{-1}\sum_{i=1}^{n}\boldsymbol{u}(\boldsymbol{A}_i)\right),$$

として考え得る．ここで $\phi(\theta)$ は実直線上の一つの連結領域上で定義される θ の連続かつ狭義単調関数，$\phi_{\boldsymbol{U}}^{-1}$ は $\phi(\theta)$ の逆関数，$\boldsymbol{A}_i (i = 1, \ldots, n)$ はランダム行列 $\boldsymbol{A}$ に対する n 個の観測行列を表す．与えられた $\tilde{\boldsymbol{U}}$ に対して方程式 $\partial(-\Psi(\boldsymbol{\beta}))/\partial\boldsymbol{\beta} = \tilde{\boldsymbol{U}}$ は唯一の解 $\tilde{\boldsymbol{\beta}} = \boldsymbol{\beta}(\tilde{\boldsymbol{U}})$ を持つから，我々は P の発展モデルとして

$$dQ_{\beta}^* = \frac{\mathrm{etr}\{-\tilde{\boldsymbol{\beta}}^t\boldsymbol{u}(\boldsymbol{A})\}}{\boldsymbol{Z}(\tilde{\boldsymbol{\beta}})}dP(\boldsymbol{A}),\quad \boldsymbol{A} \in R^{n\times n}$$

の採用が可能である．

6. 分布の逐次発展

本章では Legendre 変換を逐次適切に用いることにより基礎の統計モデル P から出発して各段階で最適な発展モデルの列 $\{Q_i^*\}$ $(i = 1, 2, \ldots)$ を構成し，第 m 段階で許容誤差の範囲内で発展モデル Q に原理的に到達出来ることを示す．この為には命題 3.2 と定理 5.1 を系統的に利用して $I(Q\,;Q_m^*) \to 0 (m \to \infty)$ を示せばよい．Cantor の区間縮小列の構成方法との類推で，それ以前のものに単調に含まれる確率分布の部分集合族の列を考える．今 $\mathbb{F}_k$ を独立な k 個の平均値の条件を満たす確率分布族とする：

$$\mathbb{F}_k = \left\{Q_i : \int_{R^{n\times n}}\boldsymbol{u}_i(\boldsymbol{A})\,dQ_i(\boldsymbol{A}) = \boldsymbol{U}_i\ \ (i = 1, 2, \ldots, k)\right\},\quad (k = 1, 2, \ldots). \tag{6.1}$$

ここに $\boldsymbol{u}_i(\boldsymbol{A})$ は，各 i に対して第 3 章で導入した $\boldsymbol{u}(\boldsymbol{A})$ と同等な意味を持つ与えられた可測関数とする．但し以下で $\mathscr{B}[\boldsymbol{u}_i(\boldsymbol{A})]$ が十分部分 Borel 集合体となる場合は結果が自明となるので除外する（cf. 注 3.2）．明らかに $\mathbb{F}_1 \supset \mathbb{F}_2 \supset \cdots \supset \mathbb{F}_m \supset \cdots$ である．この縮小集合列に対応して内部生成情報量の単調減少列を構成出来ればよい（cf. Kullback（1968））．まず $\mathbb{F}_1$ の条件の下，上記の二定理から情報収支関係を満たす確率分布 Q_1^* が存在して

$$I(Q\,;P) = I(Q\,;Q_1^*) + I(Q_1^*\,;P),\quad Q_1^* \in \mathbb{F}_1 \tag{6.2}$$

が成立する．この時命題 3.1 および関連する式から

$$dQ_1^*(\boldsymbol{A}) = \frac{\mathrm{etr}\{-\boldsymbol{\beta}_1^t\boldsymbol{u}_1(\boldsymbol{A})\}}{Z_1(\boldsymbol{\beta}_1)}dP(\boldsymbol{A}) = \exp[-\mathrm{tr}\{\boldsymbol{\beta}_1^t\boldsymbol{u}_1(\boldsymbol{A})\} - \Psi_1(\boldsymbol{\beta}_1)]\cdot dP(\boldsymbol{A}),$$

$$Z_1(\boldsymbol{\beta}_1) = \int_{R^{n\times n}}\mathrm{etr}\{-\boldsymbol{\beta}_1^t\boldsymbol{u}_1(\boldsymbol{A})\}\,dP(\boldsymbol{A}),\quad \Psi_1(\boldsymbol{\beta}_1) = \ln Z_1(\boldsymbol{\beta}_1),\quad \boldsymbol{A} \in R^{n\times n},$$

$$U_1(\beta_1) = -\frac{\partial}{\partial \beta_1}\ln Z_1(\beta_1) = -\frac{\partial}{\partial \beta_1}\Psi_1(\beta_1),$$
$$I(Q_1^*;P) = -\mathrm{tr}\{\beta_1^t U_1(\beta_1)\} - \Psi_1(\beta_1).$$

次に Q_1^* を新たな基礎モデルと見てこれを更新することを考えよう．(6.2) 式において Q_1^* が情報源という拘束的立場にあるのは第1項の$\mathbb{F}_1$での内部生成情報量である．$\mathbb{F}_1$の場合と同様に$\mathbb{F}_2$に於いて，上記の内部生成情報量を$\mathbb{F}_2$で保存する Q_2^* が存在して

(6.3) $$I(Q;Q_1^*) = I(Q;Q_2^*) + I(Q_2^*;Q_1^*), \quad Q_2^* \in \mathbb{F}_2.$$

ここで

$$dQ_2^*(A) = \frac{\mathrm{etr}\{-\beta_2^t u_2(A)\}}{Z_2(\beta_2)} dQ_1^*(A) = \frac{\mathrm{etr}[-\{\beta_1^t u_1(A) + \beta_2^t u_2(A)\}]}{Z_1(\beta_1)Z_2(\beta_2)} dP(A)$$
$$= \exp[-\mathrm{tr}\{\beta_1^t u_1(A) + \beta_2^t u_2(A)\} - \{\Psi_1(\beta_1) + \Psi_2(\beta_2)\}] \cdot dP(A),$$

$$Z_2(\beta_2) = \int_{R^{n\times n}} \mathrm{etr}\{-\beta_2^t u_2(A)\} dQ_1^*(A), \quad \Psi_2(\beta_2) = \ln Z_2(\beta_2), \quad A \in R^{n\times n},$$

$$U_2(\beta_2) = -\frac{\partial}{\partial \beta_2}\ln Z_2(\beta_2) = -\frac{\partial}{\partial \beta_2}\Psi_2(\beta_2),$$

$$I(Q_2^*;Q_1^*) = -\mathrm{tr}\{\beta_2^t U_2(\beta_2)\} - \Psi_2(\beta_2).$$

同様に$\mathbb{F}_3$に於いて

(6.4) $$I(Q;Q_2^*) = I(Q;Q_3^*) + I(Q_3^*;Q_2^*), \quad Q_3^* \in \mathbb{F}_3.$$

以下，逐次同様にして

....

(6.5) $$I(Q;Q_{m-1}^*) = I(Q;Q_m^*) + I(Q_m^*;Q_{m-1}^*), \quad Q_m^* \in \mathbb{F}_m.$$

ここで

$$dQ_m^*(A) = \frac{\mathrm{etr}\{-\beta_m^t u_m(A)\}}{Z_m(\beta_m)} dQ_{m-1}^*(A)$$
$$= \mathrm{etr}\left\{-\sum_{i=1}^m \beta_i^t u_i(A)\right\} \Big/ \prod_{i=1}^m Z_i(\beta_i) \cdot dP(A)$$
$$= \exp\left[-\mathrm{tr}\left\{\sum_{i=1}^m \beta_i^t u_i(A)\right\} - \left\{\sum_{i=1}^m \Psi_i(\beta_i)\right\}\right] \cdot dP(A), \quad A \in R^{n\times n},$$

$$Z_m(\beta_m) = \int_{R^{n\times n}} \mathrm{etr}\{-\beta_m^t u_m(A)\} dQ_{m-1}^*(A)$$
$$= \int_{R^{n\times n}} \mathrm{etr}\left\{-\sum_{i=1}^m \beta_i^t u_i(\beta_i)\right\} \Big/ \prod_{i=1}^{m-1} Z_i(\beta_i) \cdot dP(A), \quad A \in R^{n\times n},$$

$$U_m(\beta_m) = -\frac{\partial}{\partial \beta_m}\ln Z_m(\beta_m) = -\frac{\partial}{\partial \beta_m}\Psi_m(\beta_m),$$

$$I(Q_m^*;Q_{m-1}^*) = -\mathrm{tr}\,\{\boldsymbol{\beta}_m^t \boldsymbol{U}_m(\boldsymbol{\beta}_m)\} - \{\Psi_m(\boldsymbol{\beta}_m)\}$$

と表せる．上記の一連の逐次情報量収支の式から次の収支関係が直ちに従う：

(6.6) $$I(Q;P) = I(Q_1^*;P) + I(Q_2^*;Q_1^*) + I(Q_3^*;Q_2^*) + \cdots + I(Q_m^*;Q_{m-1}^*) + I(Q;Q_m^*).$$

また，それらの構成方法から明らかなように，内部生成情報量の単調減少列を得る：

(6.7) $$I(Q;Q_1^*) \geq I(Q;Q_2^*) \geq I(Q;Q_3^*) \geq \cdots \geq I(Q;Q_m^*) \geq 0.$$

よって，内部生成情報量に関し $\lim_{m\to\infty} I(Q\,;\,Q_m^*) = 0$ が成立する．即ち m が十分に大きければ $Q \cong Q_m^*[\mu]$．この時

(6.8) $$I(Q;P) \cong I(Q_1^*;P) + I(Q_2^*;Q_1^*) + I(Q_3^*;Q_2^*) + \cdots + I(Q_m^*;Q_{m-1}^*)$$

と考えてよい．移動情報量についても $\lim_{m\to\infty} I(Q_m^*\,;\,Q_{m-1}^*) = 0$ が成立するから，基礎モデル P からスタートして上記の手続きで Legendre 変換を逐次適用，モデルの最適更新列

(6.9) $$P \to Q_1^* \to Q_2^* \to Q_3^* \to \cdots \to Q_{m-1}^* \to Q_m^*$$

を構成すれば，十分大きな m に対して Q_m^* を適切な発展モデルとして採用可能となる．なお，移動情報量列は内部生成情報列の様に単調減少性を持つとは限らない．しかし上記の極限的性

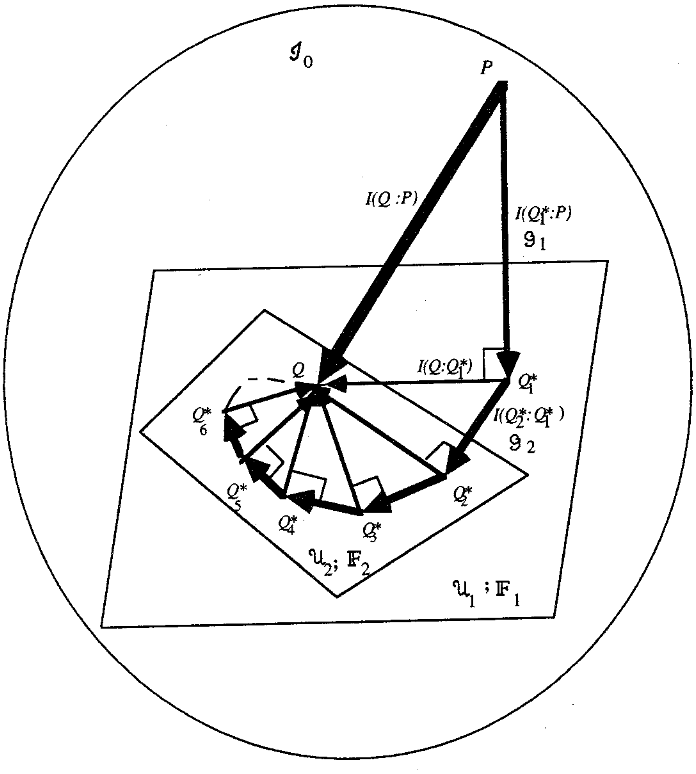

図1. 逐次 Legendre 変換によるモデルの更新.

質から，正の微小振動が続いても大局的には消滅してしまう．よって，Occamの思考節減の原理に従ってこの更新プロセスを打ち切る時点mの値としては，十分小さい$\varepsilon>0$を設定された許容誤差として，初めて条件$I(Q_m^*;Q_{m-1}^*)\leq\varepsilon$を満たす$m_0$を採択する．

注 6.1. 上の手続きを幾何学的に解釈すると次の様になる．Pから部分確率分布族$\mathbb{F}_1$に対応する部分σ–集合体の平均値の束縛条件 (6.1) を満たす被拘束超平面への射影を行ない最適更新モデルQ_1^*を得る．次にこのQ_1^*から部分確率分布族$\mathbb{F}_2$に対応する部分σ–集合体での平均束縛条件 (6.1) を満たす被拘束超平面への射影を行ない最適更新モデルQ_2^*を得る．以下同様な手続きで第mステップのQ_m^*を得る (cf. 図1)．このことから，注3.1に述べた意味での直交条件の逐次適用が**最小内部生成情報の原理** (principle of minimum internal produced information) と呼び得るものと同等な役割をしている．

注 6.2. Legendre変換を逐次使用する本章のアルゴリズムは線形計画法に於けるKarmarkar (1984) の内点法と密接に関係している．後者は離散型確率分布の最適な更新プロセスをEuclid空間で扱っていると見做せる．そこで扱われているポテンシャル関数と射影変換は本稿の場合のLegendre変換に相当する．

注 6.3. $\mathscr{B}_k=\sigma(\boldsymbol{u}_1(\boldsymbol{A}),\boldsymbol{u}_2(\boldsymbol{A}),\ldots,\boldsymbol{u}_k(\boldsymbol{A}))$を確率変数$\boldsymbol{u}_1(\boldsymbol{A}),\boldsymbol{u}_2(\boldsymbol{A}),\ldots,\boldsymbol{u}_k(\boldsymbol{A})$によって生成される$\boldsymbol{B}^{n\times n}$の部分Borel集合体とする．$\boldsymbol{u}_i(\boldsymbol{A})$ $(i=1,\ldots,k)$ の定義から $\{E(\boldsymbol{u}\mid\mathscr{B}_k),\ \mathscr{B}_k,k\in I\}$, $I=\{1,2,\ldots,+\infty\}$ はマルチンゲールを構成している (cf. Galambos (1988))．このことを考慮して我々は上記のアルゴリズムに於ける収束プロセスを取り扱える．

7. あとがき

本稿ではKullback-Leiblerの情報量を用いて，与えられた多変量統計基礎モデルがある一般的な条件のもとで移動情報最大 (or 内部生成情報最小) の意味で最適な統計モデルへ発展する可能性を正準情報量規準の形で考察した．このことを考察可能にしていることの一つはK–L情報量が負のエントロピーであることにある．この事実はBoltzmannにその起源を見ることができ，現代の統計学の研究者達も (時に符号を誤って) 言及する基本事項である．しかしながらこの重要な関係は統計学に十分に理解されて取り込まれていない．この部分に主として関係する統計理論は最尤法とその精緻な修正の理論展開であるが，研究の進展につれて統計理論そのものと関連諸科学との交流が薄れてしまったように思える．この点を反省して，筆者は本稿で，Kullback (1952) の情報不等式および情報収支の関係式の中に潜んでいた重大な結果に注目し，従来の統計理論よりもっと基礎的な立場から表題の課題に取り組んだ．その際，統計学に関連する諸科学で現われる双対問題の解析に重要な意味を持つLegendre変換およびPrigogineの内部生成エントロピーの類推として内部生成情報量が非常に有用な役割を演じることが分かった．なお，Legendre変換はAmari (1985) の情報幾何学による統計理論にも重要な役割を演じるものとして登場している．本稿での分布の発展という課題には，熱力学的な解析的手法の中でのこの変換の活用を行なうことも関連諸科学との相互理解に役立つと考えそのような接近方法を取った．情報幾何学の理解の上にその積極的利用を行なえば，本問題に対するより系統的で理解しやすいアプローチが提案できる可能性もある．また逆に，ここで得た知見が情報幾何の分野に新しい解釈をもたらすかもしれない．例えば本稿で提案した移動情報と内部生成情報の関係は，双対平坦な情報空間に於ける互いに双対な情報量あるいは情報測地線に

物理的意味を付与したとも解釈できる.

本稿執筆の過程で，統計の基礎理論が驚くほど奇麗に Legendre 変換に乗って来るということを実感した. このことは逆に統計理論が関連諸科学から共通の理解が得られる土壌を潜在的に有していると言えよう. 様々な分野でのランダムな要因に起因する発展にも本稿の結果は部分的に当て嵌まるかもしれない.

最後に，本稿では基礎モデルを与えられたものとして議論を開始したが，その様な状況は現実にはむしろ特別な場合であり，実際の多くの場面では基礎モデルをどう設定するかが問題である. 近代推測統計学においては，頻度論者は母集団分布関数を，ベイジアンは事前分布の存在を仮定して統計理論を展開するが，このことは何に根拠をもってなされるのかに疑問を持つ関連分野の研究者は決して少なくない. この疑問に答えるためにも統計家は基礎モデルの分布の特定化に真剣に取り組む必要があろう. これは分布の起源を探るという内容を持つ問題に通じる. しかし，この大きな課題については稿を改めて考察する (cf. Matsunawa (1994)).

謝　辞

本稿に対し著者の不完全な記述を正す貴重なコメントと文献の紹介を頂いた査読者に深く感謝します.

参　考　文　献

Amari, S. (1985). Differential-geometrical methods in statistics, *Lecture Notes in Statist.*, **28**, Springer, New York.

Arnold, V. I. (1980). *Geometrical Methods in the Theory of Ordinary Differential Equations*, 2nd ed., Springer, New York.

Barndorff-Nielsen, O. E. and Cox, D. R. (1989). *Asymptotic Techniques for Use in Statistics*, Chapman and Hall, London.

Boltzmann, L. (1887). Über die Beziehung zwischen dem zweiten Hauptsatze der mechanishen Wärmetheorie und der Wahrscheinlichkeitsrechnung respektive den Sätzen über das Wärmegleichgewicht, *Wiener Berichte*, **76**, 373-435.

Brillouin, L. (1956). *Science and Information Theory*, Academic Press, New York.

Brown, L. D. (1986). *Fundamentals of Statistical Exponential Families with Applications in Statistical Decision Theory*, IMS Lecture Notes-Monograph Series, Vol. 9, IMS, Hayward, California.

Csiszár, I. (1975). I-divergence geometry of probability distributions and minimization problems, *Ann. Probab.*, **3**, 146-158.

Deuscel, J.-D. and Strook, D. W. (1989). *Large Deviations*, Academic Press, Boston.

Dobrushin, R. L. (1960). Passage to the limit under the information and entropy signs, *Theory Probab. Appl.*, **5**, 25-32.

Galambos, J. (1988). *Advanced Probability Theory*, Marcel Dekker, New York.

Ghurye, S. G. (1968). Information and sufficient sub-fields, *Ann. Math. Statist.*, **38**, 2056-2066.

Halmos, P. R. and Savage, L. J. (1949). Application of the Radon-Nikodym theorem to the theory of sufficient statistics, *Ann. Math. Statist.*, **20**, 225-241.

Hiraoka, M. and Tanaka, M. (1994). *Transport Phenomena*, Asakura, Tokyo (in Japanese).

Kallianpur, C. (1960). On the amount of information contained in a σ-field, *Contributions to Probability and Statistics —Essays in Honor of Harold Hotelling—*, 265-273, Stanford University Press, California.

Kapur, J. N. (1989). *Maximum-entropy Models in Science and Engineering*, Wiley Eastern, New Delhi.

Kapur, J. N. and Kesavan, H. K. (1992). *Entropy Optimization Principles with Applications*, Academic Press, Boston.

Karmarkar, N. (1984). A new polynomial-time algorithm for linear programing, *Combinatorica*, **4**, 373-395.

Kubo, R. (1952). *Statistical Mechanics*, Kyoritsu Shuppan, Tokyo (in Japanese).

Kullback, S. (1952). Certain inequalities in information theory and the Cramér-Rao inequality, *Ann. Math. Statist.*, **25**, 745-751.

Kullback, S. (1959). *Information Theory and Statistics*, Wiley, New York.

Kullback, S. (1968). Probability densities with given marginals, *Ann. Math. Statist.*, **39**, 1236-1243.

Kullback, S. and Leibler, R. A. (1951). On information and sufficiency, *Ann. Math. Statist.*, **22**, 79-86.

Matsunawa, T. (1994). Origin of distributions—Nonparametric statistical uncertainty relation and a statistical fundamental equation—, *Proc. Inst. Statist. Math.*, **42**, 197-214 (in Japanese).

Prigogine, I. and Defay, R. (1954). *Chemical Thermodynamics*, Longmanns Green, London.

Rockafellar, R. T. (1970). *Convex Analysis*, Princeton University Press, New Jersey.

Proceedings of the Institute of Statistical Mathematics Vol. 43, No. 2, 293–311 (1995)

Development of Distributions
— The Legendre Transformation and Canonical Information Criteria —

Tadashi Matsunawa

(The Institute of Statistical Mathematics)

Developments of given basic statistical models are investigated by the Legendre transformation with respect to their statistical characteristic functions. Concerning this it is shown that the minimal K–L information number defines the canonical probability distribution to a given basic positive measurable function. The fundamental information inequality of Kullback and the lower semi-continuity property of the information are also presented to the case of multivariate distributions. Further, it is proved that the minimal information is obtained by the multivariate Legendre transformation. Analogous and dual relations between the statistical fundamental theory and the thermodynamic theory are verified through the transformation. Based on those results canonical information criteria in multivariate cases are discussed. An algorithm resorting to successive use of the Legendre transformations to search an optimum statistical model within a certain admissible error bound is considered. The corresponding geometrical interpretation to the iterative procedure is shown and the connection between our algorithm and Karmarkar's one in the theory of linear programing is noticed.

Key words : K–L information, information inequality, canonical distribution, cumulative generating function, statistical characteristic function, statistical entropy function, multivariate Legendre transformation, canonical information criteria, successive development of distributions, Karmarkar's algorithm, martingale convergence.