

量子状態推定の漸近理論について

長岡浩司 (Hiroshi Nagaoka)

電気通信大学情報システム学研究科

1 はじめに

量子力学に従うある物理系に対し何らかの測定を行なったとき、どのような測定結果が得られるかは、一般には確率的にしか予言出来ない。この測定結果 x に関する確率分布 $P(dx)$ は、行なった測定の種類 Π とともに、測定対象の物理系の（測定直前の）状態 ρ に依存して $P(dx) = P(dx|\rho, \Pi)$ の様に定まる。量子力学の最もポピュラーな解釈では、状態とは物理系がどのように「用意」(preparation)されたかということに対応して定まるものとされ、数学的には、波動関数や密度作用素、より一般的には物理量の成す作用素代数上の期待値汎関数によって表される。

本稿では、真の状態があるパラメトリックな集合（モデル） $M = \{\rho_\theta | \theta \in \Theta\}$ ($\Theta \subset \mathbf{R}^d$) に属することだけが分かっているという想定のもとで、未知パラメータ θ の値を観測データから推定する「量子推定問題」について述べる。ここで、もしモデル M とともに測定 Π もあらかじめ指定されているとすると、確率分布 $P_\theta(dx) \stackrel{\text{def}}{=} P(dx|\rho_\theta, \Pi)$ に従うデータ x を見て θ を推定するという単なるパラメータ推定問題になってしまうのだが、量子推定の場合には、指定されるのはモデル M のみであり、測定 Π については θ の推定という目的のために最も適したものを求めることが要請される。こうした問題の組織的研究は、光通信の理論的研究に端を発して、1960 年代後半 C. W. Helstrom によって創められた。

後述の様に、量子推定の数学的な問題設定は、基本的には通常の統計的パラメータ推定（以下では「古典的」と形容する）の問題設定に倣った形で定式化される。例えば、量子推定においても不偏性や一致性という条件を導入することができ、それらの条件のもとで平均二乗誤差について議論することが可能である。さらに、そうした問題に対するアプローチの方法についても、Cramér-Rao 不等式に基づいた古典論的な理論展開が試みられている。しかし、そうした方法論に従って議論を進めていくとき、古典論には見られなかった多くの数学的困難が現れる。特に、パラメータの次元 d が2以上の場合には、Cramér-Rao 不等式の量子版の理論が未完成なため、任意のモデルに適用可能な一般論はほとんど存在していないという状況にある。その中でこれまでに最も成功した例は、Yuen & Lax, Helstrom, Holevo 等による量子ガウス状態の推定問題の解決であろう ([3], [1], [2])。ここでは、右対数微分という概念に基づいた Cramér-Rao 不等式の一つの量子版が本質的な役割を果たす。この問題は、古典論で言えば、分散共分散行列が既知の場合の多次元正規分布における期待値パラメータの不偏推定問題に相当する。多次元量子推定理論の現状は、この様な「初歩的」なレベルにとどまっているのである。

一方、1次元パラメータ $d = 1$ の場合には、対称対数微分という概念に基づいた Cramér-Rao 不等式の理論が一応完成しており、一般論への見通しはずっと明るい。しかしそれでも、古典論における「1次の漸近有効性」をめぐる状況が量子推定ではどうなっているのか、といった基本的な事柄でさえ、完全には解明されていない。これは、古典論における漸近正規性や超有効性などに端を発した様々な数学的議論の量子版について述べているのではない。R. A. Fisher のヴィジョンに相当する統一的な方向性そのものが、量子推定理論では未だ確立されていないのである。

以下では、量子推定の定式化および Cramér-Rao 不等式や漸近有効性に関するいくつかの数学的困難について解説する。ただし、議論が必要以上に煩雑になるのを避けるために、関数解析的な事項や各種の正則条件等についてはあまりこだわらないことにする。

2 量子系の状態と測定

量子力学では、対象とする物理系を記述する際に、ヒルベルト空間とその上の作用素という数学的な言葉を用いる。 \mathcal{H} を可分な複素ヒルベルト空間とし、その内積を $\langle \psi | \varphi \rangle$ ($\psi, \varphi \in \mathcal{H}$) と表す。ここで $\langle \psi | \varphi \rangle$ は φ に関しては線形、 ψ に関しては共役線形 ($\langle c\psi | \varphi \rangle = \bar{c} \langle \psi | \varphi \rangle$, $c \in \mathbb{C}$, \bar{c} は c の共役) とする。 \mathcal{H} 上の線形作用素 $\rho: \mathcal{H} \rightarrow \mathcal{H}$ が $\rho = \rho^* \geq 0$ (エルミート非負定値) かつ $\text{Tr} \rho = 1$ (Tr はトレース) を満たすとき、 ρ を密度作用素と呼ぶ。 \mathcal{H} で記述される量子系を考えると、その状態は一般に \mathcal{H} 上の密度作用素で表される。以下では密度作用素と状態を同一視する。状態 ρ は、 $\text{rank} \rho = 1$ のとき純粋状態、それ以外のとき混合状態と呼ばれる。

この量子系に対し、ある測定を行うとする。測定値はある集合 \mathcal{X} の要素として得られるとし、また測定結果に関する事象は \mathcal{X} 上の σ -加法族 \mathcal{B} の要素として表されるとしよう。状態 ρ におかれた系に対してこの測定を行なったとき、測定結果が $B (\in \mathcal{B})$ に属する確率は、一般に次の形に表すことができる。

$$P(B | \rho, \Pi) = \text{Tr} [\rho \Pi(B)] \quad (1)$$

ここで、 Π は \mathcal{B} の各要素 B に対して \mathcal{H} 上の作用素 $\Pi(B)$ を対応させる写像 $\Pi: \mathcal{B} \rightarrow \Pi(B)$ で、次の条件を満たす。

$$(i) \quad \forall B \in \mathcal{B}; \quad \Pi(B) = \Pi(B)^* \geq 0 \quad (\text{エルミート非負定値})$$

$$(ii) \quad \Pi(\phi) = 0, \quad \Pi(\mathbb{R}^p) = I \quad (\text{恒等作用素})$$

$$(iii) \quad B_i \cap B_j = \phi \quad (\forall i \neq \forall j) \text{ を満たす高々可算個の}$$

$$\text{集合列 } \{B_j\} \subset \mathcal{B} \text{ に対し、} \Pi\left(\bigcup_j B_j\right) = \sum_j \Pi(B_j)$$

(i)-(iii) は、 $\Pi(B)$ が作用素ではなくスカラーであるとするれば、 $(\mathcal{X}, \mathcal{B})$ 上の確率測度の定義そのものになるので、この様な Π は「確率作用素値測度」(probability operator-valued

measure) と呼ばれることがある。条件 (i)-(iii) および密度作用素の定義のもとで、式 (1) の $P(\cdot | \rho, \Pi)$ が $(\mathcal{X}, \mathcal{B})$ 上の確率測度を成すことは、簡単に確かめることができる。式 (1) を

$$P(dx | \rho, \Pi) = \text{Tr} [\rho \Pi(dx)] \quad (2)$$

と書くこともある。

上記の Π は、状態 ρ には依らず、行なう測定の種類だけから定まる。量子力学的測定の持つ二つの側面、「測定結果の統計性」と「測定による状態の変化」のうち、前者の側面はこの Π によって完全に規定される。後者の側面は本稿の内容と直接関係しないので、以下では、 Π を測定そのものと同一視してしまうことにする。

ここで、通常 (古典的な) 確率論との関係を見ておこう。まず、ヒルベルト空間 \mathcal{H} を与えるということ、測度空間 $(\Omega, \mathcal{F}, \mu)$ を与えることと対応させて考える。そして、 \mathcal{H} 上の作用素を (Ω, \mathcal{F}) 上の可測関数に、トレース Tr を Ω 上の積分 $\int d\mu$ にそれぞれ対応させてみる。このとき、 \mathcal{H} 上の状態 ρ は $(\Omega, \mathcal{F}, \mu)$ 上の確率密度関数 p に自然に対応することがわかる。ここで、可測空間 $(\mathcal{X}, \mathcal{B})$ に値をとる統計量 (確率変数) について考えよう。これは通常、 (Ω, \mathcal{F}) から $(\mathcal{X}, \mathcal{B})$ への可測な写像 $X: \Omega \rightarrow \mathcal{X}$ として表されるが、ここでは少し一般化して、データ ω が得られたときの値 x は (条件付き) 確率分布 $Q(dx | \omega)$ にしたがってランダムに決まる、という状況を考える。データ ω から deterministic に定まる統計量 (語弊があるかもしれないが、以下 deterministic な統計量と呼ぶ) X は、 $Q_X(dx | \omega) = \delta(x - X(\omega))dx$ という特別な場合であると考えればよい。このとき、分布 (密度関数) p のもとでの統計量 Q の振る舞いは

$$P(dx | p, Q) = \int p(\omega) Q(dx | \omega) d\mu(\omega)$$

によって表され、式 (2) と対応した形になる。つまり、量子測定 Π は条件付分布 Q に対応した概念であることがわかる。

古典的な場合の deterministic な統計量 $Q = Q_X$ に相当する量子的概念としては、 \mathcal{X} -値測定の条件 (i)-(iii) に加えて

$$(iv) \quad \forall B_1, B_2 \in \mathcal{B}; \quad \Pi(B_1 \cap B_2) = \Pi(B_1)\Pi(B_2)$$

を満たす様な Π を考えることができる。この様な Π は単純測定と呼ばれ、余分な不確定性を持たない「純粋な」測定を表していると考えられる。(これに対し (i)-(iii) のみを満たす Π を一般化測定と呼ぶこともある。) 特に実数値の単純測定 (オブザーバブルと呼ぶこともある) Π は、スペクトル積分 (固有値分解の一般化)

$$X = \int x \Pi(dx) \quad (3)$$

によって、自己共役作用素 X と 1対1 に対応する。この場合、状態 ρ のもとでの測定結果の期待値および分散は

$$\begin{aligned} \mu &= \int x \operatorname{Tr} [\rho \Pi(dx)] = \operatorname{Tr} [\rho X] \\ \sigma^2 &= \int (x - \mu)^2 \operatorname{Tr} [\rho \Pi(dx)] = \operatorname{Tr} [\rho (X - \mu I)^2] \end{aligned}$$

(I は恒等作用素) と表される。作用素 X は、古典的な場合の実数値確率変数 $X : \Omega \rightarrow \mathbf{R}$ に対応する概念と考えられる。

古典的な場合に deterministic な統計量を「ランダム化」して条件付き分布 Q が得られた様に、単純測定をランダム化すれば一般化測定が得られる。しかし、すべての一般化測定が単純測定のランダム化によって得られるわけではない。また、古典推定理論では deterministic な推定量に話を限ってもほとんど一般性を失わなかったが、未知パラメータが複数の場合の量子推定では、後述の様に一般化測定を導入することが本質的に必要になる。

3 量子状態推定の数学的定式化

正の密度作用素の全体を

$$S = S(\mathcal{H}) \stackrel{\text{def}}{=} \{\rho \mid \rho = \rho^* > 0, \operatorname{Tr} \rho = 1\}$$

とおき、その部分集合として、状態の候補の集合

$$M = \{\rho_\theta \mid \theta = (\theta^1, \dots, \theta^d) \in \Theta\} \subset S$$

が与えられたとする。ただし Θ は \mathbf{R}^d の開集合であり、対応 $\theta \mapsto \rho_\theta$ は 1対1かつ適当な意味で滑らかであるとする。以下では M をモデルと呼ぶ。なお、ここで仮定した状態の

正值性 $\rho_\theta > 0$ は、通常の統計的推定理論でしばしば仮定される確率密度関数の正值性（分布の台がパラメータの値に寄らず一定）と同様の、理論展開を容易にするための正則条件の一種である。この条件によって、例えば純粋状態の推定などは排除されるが、以下の議論のある部分については純粋状態推定にも適用できる。

真の状態が M のある要素 ρ_θ によって表されることは分かっているが θ の値が未知であるという状況の下で、その値を如何にして推定するかという問題について考えよう。この場合、推定量（推定方法）とは Θ に値をとる測定のことであると考えることができる。ここでは（やはり理論展開を容易にするために）条件を少し緩めて、推定量とは R^d に値をとる測定、すなわち (R^d, \mathcal{B}) 上の確率作用素値測度のことであると定義してしまうことにする。また、ある集合 \mathcal{X} に値をとる測定 $\tilde{\Pi}(dx)$ を行ない、その測定結果 $x \in \mathcal{X}$ に基づいて推定値 $t = T(x) \in R^d$ を求める場合を考えると、これは $\Pi(dt) = \tilde{\Pi}(T^{-1}(dt))$ なる推定量 Π を定める。このような Π を、測定 $\tilde{\Pi}$ に基づいた推定量と呼ぶ。

推定量 Π を用いたとき、状態 ρ のもとでの推定値の確率分布は

$$P(dt | \rho, \Pi) = \text{Tr} [\rho \Pi(dt)] \quad (4)$$

で与えられる。特に $\rho = \rho_\theta$ の場合には、 $P_\theta(dt | \Pi) = P(dt | \rho_\theta, \Pi)$ と書くことにする。

モデル $M = \{\rho_\theta\}$ と推定量 Π が与えられれば、推定値の分布 $P_\theta(dt | \Pi)$ が定まるので、古典的な推定理論と同様な種々の問題設定が可能になる。例えば、loss function L および prior W を導入することによって、Bayes risk

$$R(\Pi) = \int L(t, \theta) P_\theta(dt | \Pi) W(d\theta)$$

を考え、これを最小にする様な推定量 Π を求めるという定式化もできる。しかし以下では、不偏推定と分散に関する議論についてのみ述べることにする。

4 不偏推定と Cramér-Rao 不等式

$\forall \theta \in \Theta$ に対し

$$\int t^i P_\theta(dt | \Pi) = \theta^i \quad (\forall i) \quad (5)$$

が成り立つとき、 Π は不偏推定量であると言う。この場合、推定値の分散共分散は

$$v_{\theta}^{ij} = \int (t^i - \theta^i)(t^j - \theta^j) P_{\theta}(dt | \Pi) \quad (6)$$

を成分とする $d \times d$ 行列 $V_{\theta}[\Pi]$ で表される。

ここで、 $\forall i, \forall \theta$ に対し

$$\frac{\partial}{\partial \theta^i} \rho_{\theta} = \frac{1}{2}(\rho_{\theta} L_{\theta,i} + L_{\theta,i} \rho_{\theta}) \quad (7)$$

を満たす自己共役作用素 $L_{\theta,i} = L_{\theta,i}^*$ をとり、

$$J_{\theta,ij} \stackrel{\text{def}}{=} \text{Re Tr} [\rho_{\theta} L_{\theta,i} L_{\theta,j}] \quad (8)$$

(Re は実部) を成分とする $d \times d$ 行列 J_{θ} を定義すると、不偏な Π に対して

$$V_{\theta}[\Pi] \geq J_{\theta}^{-1} \quad (9)$$

(「左辺 - 右辺」が非負定値) が成り立つことが示せる (Helstrom (1967); [1], [2] 参照)。作用素 $L_{\theta,i}$ はモデル $M = \{\rho_{\theta}\}$ の対称対数微分と呼ばれる。

モデル M を確率密度関数の成すモデル $\{p_{\theta}\}$ に対応させてみると、 $L_{\theta,i}$ は $\frac{\partial}{\partial \theta^i} \log p_{\theta}$ に、 J_{θ} は Fisher 情報行列に、式 (9) は Cramér-Rao 不等式 (以下 CR 不等式と略記する) にそれぞれ対応することが分かる。しかし、古典的な CR 不等式と量子 CR 不等式 (9) との間には、実は本質的な相違がある。以下、それを説明しよう。

推定量 Π が不偏ならば、 $\forall \theta \in \Theta$ に対し式 (5) が成り立つので、それを微分した式

$$\int t^i \frac{\partial}{\partial \theta^j} P_{\theta}(dt | \Pi) = \delta_j^i \quad (\forall i, \forall j) \quad (10)$$

もまた $\forall \theta \in \Theta$ に対し成り立つ (微分と積分の順序は交換できるものと仮定する)。ここで不偏性の条件をゆるめ、 Θ の一点 θ において式 (5) (10) が成り立つとき、 Π は θ において局所不偏であると言うことにする。これは、点 θ において1次近似の意味で不偏性が成り立つことを意味する。点 θ における不等式 (9) は、 θ における局所不偏性だけから導かれる。このことは、古典的 CR 不等式についても同様である。古典的な場合にはさらに、CR 不等式の下限を達成する局所不偏な推定量が各点 θ 毎に常に存在する。(ただし θ に

ついて一様に下限を達成する不偏推定量（有効推定量）は、モデルが指数型分布族の場合を除いて存在しない。）この意味で、古典的な CR 不等式は最良な評価である。この最良性は、漸近的には CR 不等式の下限が常に達成可能である（i.e. 1 次の漸近有効推定量が存在する）という重要な事実と密接に関係している。

一方、量子系で $d \geq 2$ の場合、各点 θ 毎に見ても、不等式 (9) の等号を達成する局所不偏な推定量は一般には存在しない。これは J_θ の定義の仕方が悪いわけではなく、その様な達成可能な下限は原理的に存在しないのである。その背後には、複数の非可換な物理量の同時測定不可能性という所謂「不確定性原理」が関わっている。こうした困難のために、 $d \geq 2$ の量子推定では分散共分散行列 $V_\theta[\Pi]$ 自体の最適化は、ほとんどの場合不可能である。そこで、例えば適当な $d \times d$ の正定値行列 W を「重み」として与えてやり、スカラー量 $\text{tr}(WV_\theta[\Pi])$ の最小化を目指す（ tr は $d \times d$ 行列のトレースを表す；cf. Tr ）という様な方針を取らざるを得ない。この場合、 θ において局所不偏な Π に対する $\text{tr}(WV_\theta[\Pi])$ の達成可能な下限を求めることが問題になるが、これについて満足すべき一般的な結果はまだ得られていない。現状ではいくつかの下限が知られているだけである。例えば、上述の J_θ を用いた $\text{tr}(WJ_\theta^{-1})$ もその一つである。別の下限としては、右対数微分という概念を用いたものがよく知られている。Yuen & Lax [3] は、量子ガウス状態の期待値パラメータの推定という重要な問題に対し、右対数微分による下限が達成可能であることを示すことによって最適な推定量（重み W に依存する）を求めた。

以上は $d \geq 2$ に特有の困難であり、 $d = 1$ の場合には、古典的 CR と同様に不等式 (9) は θ 毎に常に達成可能である。しかも、推定量として単純測定をとることができる。これを確かめよう。以下では、推定量 Π として R -値単純測定を考え、それを自己共役作用素 $T = \int t\Pi(dt)$ によって表すことにする。このとき、局所不偏性条件 (5) (10) は

$$\text{Tr}[\rho_\theta T] = \theta, \quad \text{Tr}\left[\left(\frac{d}{d\theta}\rho_\theta\right)T\right] = 1 \quad (11)$$

と表される。ここで θ を任意に固定し、 θ における対称対数微分 L_θ を用いて T を

$$T = \theta I + \frac{1}{J_\theta} L_\theta \quad (12)$$

と定めれば、(11) が成り立ち、かつ

$$V_\theta[T] = \text{Tr} [\rho_\theta(T - \theta I)^2] = \frac{1}{J_\theta} \quad (13)$$

となる。すなわち、 T は θ における CR 不等式 (9) の下限を達成する。本稿ではこれより以降 $d = 1$ の場合のみを扱うので、推定量はいつも自己共役作用素として表すことにする。

5 合成系と i.i.d. 条件

n 個の量子系がそれぞれヒルベルト空間 $\mathcal{H}_1, \mathcal{H}_2, \dots, \mathcal{H}_n$ によって表されるとき、それらを合成した系はテンソル積ヒルベルト空間

$$\mathcal{H}^{(n)} = \mathcal{H}_1 \otimes \mathcal{H}_2 \otimes \dots \otimes \mathcal{H}_n = \bigotimes_{i=1}^n \mathcal{H}_i$$

によって表される。したがって、その合成系の状態は $\mathcal{H}^{(n)}$ 上の密度作用素 $\rho^{(n)}$ によって表される。特に、 n 個の要素系が互いに独立に状態 $\rho_1, \rho_2, \dots, \rho_n$ におかれているという状況は、

$$\rho^{(n)} = \rho_1 \otimes \rho_2 \otimes \dots \otimes \rho_n = \bigotimes_{i=1}^n \rho_i$$

と表される。さらに

$$\mathcal{H}_1 = \mathcal{H}_2 = \dots = \mathcal{H}_n, \quad \rho_1 = \rho_2 = \dots = \rho_n \quad (14)$$

の場合を考えれば、これが古典的な i.i.d. (independent and identically distributed) に相当する状況である。

合成系 $\mathcal{H}^{(n)} = \bigotimes_{i=1}^n \mathcal{H}_i$ 上の自己共役作用素 $T^{(n)}$ は

$$T = \int t \Pi^{(n)}(dt) \quad (15)$$

とスペクトル表現される。ここで $\Pi^{(n)}$ は $\mathcal{H}^{(n)}$ 上のスペクトル測度であるが、これは一般にどのような測定を表すのだろうか。

まず、各要素系 $\mathcal{H}_1, \dots, \mathcal{H}_n$ に対し独立に (単純) 測定 $\Pi_1(dx_1), \dots, \Pi_n(dx_n)$ を行ない、測定結果 (x_1, \dots, x_n) の関数として実数 $t = t(x_1, \dots, x_n)$ を得る、という形の実数値測定を考えよう。これは、

$$T^{(n)} = \int \cdots \int t(x_1, \dots, x_n) \bigotimes_{i=1}^n \Pi_i(dx_i) \quad (16)$$

という自己共役作用素 $T^{(n)}$ によって表される。この $T^{(n)}$ のスペクトル表現 (15) は

$$\Pi^{(n)}(dt) = \int \cdots \int_{t(x_1, \dots, x_n) \in dt} \bigotimes_{i=1}^n \Pi_i(dx_i) \quad (17)$$

で与えられる。先に述べた i.i.d. 条件: $\mathcal{H}_i = \mathcal{H}$, $\rho_i = \rho$ ($\forall i$) のもとで、さらに $\Pi_i = \Pi$ ($\forall i$) も成り立つとすると、 (x_1, \dots, x_n) は分布 $P(dx) = \text{Tr}[\rho \Pi(dx)]$ に従う (古典的な意味での) i.i.d. データになる。

上記の測定をもう少し一般化すると、 $\mathcal{H}_1, \dots, \mathcal{H}_n$ に対し測定 Π_1, \dots, Π_n を順番に行なっていくときに、 i 番目の測定 Π_i をそれまでに得られたデータ (x_1, \dots, x_{i-1}) に依存して定めていく、という状況を考えることができる。これは

$$T^{(n)} = \int \cdots \int t(x_1, \dots, x_n) \bigotimes_{i=1}^n \Pi_i(dx_i | x_1, \dots, x_{i-1}) \quad (18)$$

$$\Pi^{(n)}(dt) = \int \cdots \int_{t(x_1, \dots, x_n) \in dt} \bigotimes_{i=1}^n \Pi_i(dx_i | x_1, \dots, x_{i-1}) \quad (19)$$

と表される。

$\mathcal{H}^{(n)}$ 上の実数値測定 $\Pi^{(n)}(dt)$ は、一般には式 (17) (19) の形には表せない。一般の $\mathcal{H}^{(n)}$ 上の測定は、上述の様な「各要素系に対し一つずつ測定を行なっていく」ことによっては実現できず、要素系間の何らかの量子力学的相互作用を利用することが必要となる。

6 漸近分散

\mathcal{H} 上の 1-パラメータ・モデル $M = \{\rho_\theta | \theta \in \Theta\}$ ($\Theta \subset \mathcal{R}$) に対し、その n -i.i.d. 拡張 $M^{(n)} = \{\rho_\theta^{(n)} | \theta \in \Theta\}$, $\rho_\theta^{(n)} = \bigotimes^n \rho_\theta$ を考えよう。モデル $M^{(n)}$ に関する推定量は、

$\mathcal{H}^{(n)} = \otimes^n \mathcal{H}$ 上の自己共役作用素 $T^{(n)}$ によって表され、状態 $\rho_\theta^{(n)}$ のもとでの推定値の分布は、スペクトル表現 (15) を用いて

$$P_\theta(dt | T^{(n)}) = \text{Tr} [\rho_\theta^{(n)} \Pi^{(n)}(dt)] \quad (20)$$

と書ける。 $\{\mathcal{H}^{(n)}; n = 1, 2, \dots\}$ 上の推定量の系列 $\{T^{(n)}; n = 1, 2, \dots\}$ が、 $\forall \theta \in \Theta, \forall \epsilon > 0$ に対し

$$P_\theta\{|T^{(n)} - \theta| > \epsilon\} = \int_{|t-\theta|>\epsilon} P_\theta(dt | T^{(n)}) \longrightarrow 0 \quad \text{as } n \rightarrow \infty \quad (21)$$

を満たすとき、 $\{T^{(n)}\}$ を一致推定量と呼ぶ。

モデル $M^{(n)}$ について対称対数微分に基づいた Fisher 情報量 $J_\theta^{(n)}$ を考えると、 M の Fisher 情報量 $J_\theta = J_\theta^{(1)}$ により

$$J_\theta^{(n)} = nJ_\theta \quad (22)$$

と表されることが容易に示せる。したがって、 θ において局所不偏な任意の推定量 $T^{(n)}$ に対し

$$V_\theta[T^{(n)}] = \text{Tr} [\rho_\theta^{(n)} (T^{(n)} - \theta I)^2] \geq \frac{1}{nJ_\theta} \quad (23)$$

が成り立つ。また、局所不偏性を仮定しない一般の場合にも、平均二乗誤差 V_θ (\neq 分散) に関して

$$V_\theta[T^{(n)}] \geq \left(\frac{d}{d\theta} E_\theta[T^{(n)}]\right)^2 / nJ_\theta + (E_\theta[T^{(n)}] - \theta)^2 \quad (24)$$

が成り立つ ($E_\theta[T^{(n)}] = \text{Tr} [\rho_\theta^{(n)} T^{(n)}$)。このことから、古典的な場合と同様に、一致性 (21) に加えてある種の正則条件を課すことにより、漸近的 CR 不等式

$$\lim_{n \rightarrow \infty} nV_\theta[T^{(n)}] \geq \frac{1}{J_\theta} \quad (25)$$

を示すことができる。

次に、不等式 (25) の等号を達成する様な一致推定量 $\{T^{(n)}\}$ の存在について考えてみよう。通常の (ある種の正則条件を満たす) 古典的モデルでは、最尤推定 (MLE) がこの下限を達成する推定量の一例になる。量子系では MLE の直接的な対応物は一般に存在しないが、もし \mathcal{H} 上の測定 $\Pi(dx)$ を固定して、分布 $P_\theta(dx | \Pi) = \text{Tr} [\rho_\theta \Pi(dx)]$ の成す古典的モ

デル $M(\Pi)$ を考えれば、それについては MLE を考えることができる。即ち、測定 Π を独立に n 回行ない、得られた (x_1, \dots, x_n) に基づいて、最尤推定値

$$t = t(x_1, \dots, x_n) = \arg \min_{\theta} p_{\theta}(x_1 | \Pi) \cdots p_{\theta}(x_n | \Pi), \quad \text{where } p_{\theta}(x | \Pi) \stackrel{\text{def}}{=} \frac{P_{\theta}(dx | \Pi)}{dx}$$

を求めればよい。この手続きは、式 (16) (17) 型の推定量 $T^{(n)}$ を定める。この推定量に対し、モデル $M(\Pi)$ の Fisher 情報量を $J_{\theta}[\Pi]$ とおくと、

$$\lim_{n \rightarrow \infty} nV_{\theta}[T^{(n)}] = \frac{1}{J_{\theta}[\Pi]} \quad (26)$$

が成り立つ。

古典的 Fisher 情報量 $J_{\theta}[\Pi]$ と量子的 Fisher 情報量 J_{θ} の間には、次の関係がある ([4])。

$$J_{\theta} = \max_{\Pi} J_{\theta}[\Pi] \quad (27)$$

この \max は、 θ における対称対数微分 L_{θ} のスペクトル分解 Π_{θ} (式 (12) の T のスペクトル分解でも同じ) によって達成される。そこで、任意に固定した θ については、 Π_{θ} に基づいた MLE を行なうことによって式 (25) の等号が達成できることがわかる。しかし、一般に Π_{θ} は θ に依存するので、この方法は θ の推定には使えない。

この問題を解決する一つの方策として、推定を逐次的に行ないながら、各時点での推定値を用いて次の測定を定めていく、という推定法が考えられる ([5])。まず、推定値の初期値 t_0 を任意に (あるいは適当に) 定め、測定 Π_{t_0} を行なう。得られたデータ x_1 を用いて、モデル $\{P_{\theta}(dx_1 | \Pi_{t_0})\}$ に関する最尤推定値 t_1 を求める。次に測定 Π_{t_1} を行ない、得られたデータ x_2 とし、 (x_1, x_2) に基づいてモデル $\{P_{\theta}(dx_1 | \Pi_{t_0})P_{\theta}(dx_2 | \Pi_{t_1})\}$ に関する最尤推定値 t_2 を求める。・・・測定 $\Pi_{t_{n-1}}$ を行ない、得られたデータを x_n とし、 (x_1, \dots, x_n) に基づいてモデル $\{P_{\theta}(dx_1 | \Pi_{t_0}) \cdots P_{\theta}(dx_n | \Pi_{t_{n-1}})\}$ の最尤推定値

$$t_n = \arg \min_{\theta} p_{\theta}(x_1 | \Pi_{t_0}) \cdots p_{\theta}(x_n | \Pi_{t_{n-1}})$$

を求める。この方法は、式 (18) (19) 型の推定量になる。 $n \rightarrow \infty$ において t_n が真値 θ に近づいていくなれば、 Π_{t_n} も Π_{θ} に近づいていき、 t_n は Π_{θ} に基づく MLE に近づいてい

き、したがって式 (25) の等号が達成されるだろう、と予想される。ただし、この予想の厳密な検討はまだ成されていない。なおこの問題は、2種類のパラメータ θ, ξ を持つ古典的モデル $\{P(dx|\theta, \xi)\}$ において、 θ は未知であるが ξ は観測者が自由に設定できる、という状況のもとで θ をどれくらいの精度で推定できるか、という古典的推定問題の特別な場合とみなすことができる。

7 大偏差型評価

ある種の正則条件のもとで、古典的な 1-パラメータ・モデル $\{P_\theta(dx) | \theta \in \Theta\}$ ($\Theta \subset \mathbf{R}$) の Fisher 情報量 J_θ は、Kullback divergence (Kullback-Leibler 情報量):

$$D(P \parallel Q) = \int (\log \frac{dP}{dQ}) dP \quad (28)$$

を用いて

$$J_\theta = 2 \lim_{\Delta\theta \rightarrow 0} D(P_{\theta+\Delta\theta} \parallel P_\theta) / (\Delta\theta)^2 \quad (29)$$

と表される。量子系の場合、Kullback divergence の対応物として知られているのは、次の 量子相対エントロピー (梅垣エントロピー) である ([9])。

$$D(\rho \parallel \sigma) = \text{Tr} [\rho (\log \rho - \log \sigma)] \quad (30)$$

そこで、量子モデル $\{\rho_\theta\}$ に対し

$$\tilde{J}_\theta \stackrel{\text{def}}{=} 2 \lim_{\Delta\theta \rightarrow 0} D(\rho_{\theta+\Delta\theta} \parallel \rho_\theta) / (\Delta\theta)^2 \quad (31)$$

とおけば、これも Fisher 情報量の一種の量子版とみなせる。しかし、対称対数微分による Fisher 情報量 J_θ との関係は

$$J_\theta \leq \tilde{J}_\theta \quad (32)$$

であって、等号は一般には成立しない ([6])。この事実は、推定論的にどのような意味を持つのだろうか。

古典的な確率論において kullback divergence が基本的な役割りを果たすのは、いわゆる大偏差 (large deviation) 型の問題である (Cramér, Sanov, Stein, etc.)。式 (29) を大偏差の一般理論と結びつけることにより、次の 大偏差型 CR 不等式 を示すことができる ([7], [8])。

$$\lim_{\epsilon \downarrow 0} \lim_{n \rightarrow \infty} \frac{1}{\epsilon^2 n} \log P_\theta \{|T^{(n)} - \theta| \geq \epsilon\} \geq -\frac{1}{2} J_\theta \quad (33)$$

ここで $\{T^{(n)}\}$ は任意の一致推定量である。上式の左辺は、 $\{T^{(n)}\}$ の Bahadur efficiency と呼ばれる。上式は、より基本的な不等式

$$\lim_{\epsilon \downarrow 0} \lim_{n \rightarrow \infty} \frac{1}{\epsilon^2 n} \log P_\theta \{T^{(n)} \geq \theta + \epsilon\} \geq -\frac{1}{2} J_\theta \quad (34)$$

$$\lim_{\epsilon \downarrow 0} \lim_{n \rightarrow \infty} \frac{1}{\epsilon^2 n} \log P_\theta \{T^{(n)} \leq \theta - \epsilon\} \geq -\frac{1}{2} J_\theta \quad (35)$$

から導かれる。また、 $\{T^{(n)}\}$ として最尤推定をとれば、式 (34) (35) の等号が同時に達成され、したがって式 (33) の等号も達成される。以下では、式 (34) の量子版について考える ([6])。

まず準備として、量子相対エントロピー D のいくつかの基本的性質について述べておく。ヒルベルト空間 \mathcal{H} 上の状態 ρ, σ および測定 $\Pi(dx)$ に対し、確率分布 $P(dx) = \text{Tr} [\rho \Pi(dx)]$, $Q(dx) = \text{Tr} [\sigma \Pi(dx)]$ の Kullback divergence $D(P \parallel Q)$ を $D_\Pi(\rho \parallel \sigma)$ と書くことにする。このとき

$$D_\Pi(\rho \parallel \sigma) \leq D(\rho \parallel \sigma) \quad (36)$$

が常に成り立つ。これは、相対エントロピーの 単調性 と呼ばれる性質の特別な場合に相当する ([9])。状態 ρ, σ が与えられたとき、上式の等号を達成する Π は一般には存在しない。(なお、式 (27) (29) (31) (36) より式 (32) が導かれる。)

ここで、 ρ, σ の n -i.i.d. 拡張 $\rho^{(n)} = \otimes^n \rho, \sigma^{(n)} = \otimes^n \sigma$ を考えると、

$$D(\rho^{(n)} \parallel \sigma^{(n)}) = nD(\rho \parallel \sigma) \quad (37)$$

となる。したがって、 $\mathcal{H}^{(n)} = \otimes^n \mathcal{H}$ 上の任意の測定 $\Pi^{(n)}$ に対し

$$\frac{1}{n} D_{\Pi^{(n)}}(\rho^{(n)} \parallel \sigma^{(n)}) \leq D(\rho \parallel \sigma) \quad (38)$$

が成り立つ。この等号も一般には達成不可能である。しかし、 $n \rightarrow \infty$ の極限を考えれば

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sup_{\Pi^{(n)}} D_{\Pi^{(n)}}(\rho^{(n)} \parallel \sigma^{(n)}) = D(\rho \parallel \sigma) \quad (39)$$

が常に成立する ([10])。

さて、 \mathcal{H} 上の 1-パラメータ・モデル $M = \{\rho_\theta \mid \theta \in \Theta\}$ ($\Theta \subset \mathbf{R}$) の一致推定量 $\{T^{(n)}\}$ を任意にとると、単調性 (38) により、 $\forall \theta \in \Theta, \forall \epsilon > 0$ に対し

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \log P_\theta \{T^{(n)} \geq \theta + \epsilon\} \geq -D(\rho_{\theta+\epsilon} \parallel \rho_\theta) \quad (40)$$

となることが示される ([6])。また、 $\forall \theta \in \Theta, \forall \delta > 0$ に対し、式 (39) より

$$\frac{1}{l} D_{\Pi^{(l)}}(\rho_{\theta+\epsilon}^{(l)} \parallel \rho_\theta^{(l)}) \geq D(\rho_{\theta+\epsilon} \parallel \rho_\theta) - \delta \quad (41)$$

を満たす自然数 l および $\mathcal{H}^{(l)} = \otimes^n \mathcal{H}$ 上の測定 $\Pi^{(l)}$ が存在する。この $\Pi^{(l)}$ の独立な繰り返し測定を用いた式 (16) 型の推定量 $\{T^{(n)}\}$ で、一致性および

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log P_\theta \{T^{(n)} \geq \theta + \epsilon\} = -\frac{1}{l} D_{\Pi^{(l)}}(\rho_{\theta+\epsilon}^{(l)} \parallel \rho_\theta^{(l)}) \quad (42)$$

を満たすものを構成することができる ([6])。 $\forall \delta > 0$ は任意に小さくとれるので、式 (40) (41) (42) より

$$\inf_{\{T^{(n)}\} \in \mathcal{C}} \liminf_{n \rightarrow \infty} \frac{1}{n} \log P_\theta \{T^{(n)} \geq \theta + \epsilon\} = -D(\rho_{\theta+\epsilon} \parallel \rho_\theta)$$

が成り立つ。ただし \mathcal{C} は一致推定量の全体とする。この式と (31) とから

$$\lim_{\epsilon \downarrow 0} \inf_{\{T^{(n)}\} \in \mathcal{C}} \liminf_{n \rightarrow \infty} \frac{1}{\epsilon^2 n} \log P_\theta \{T^{(n)} \geq \theta + \epsilon\} = -\frac{1}{2} \tilde{J}_\theta \quad (43)$$

が得られる。

式 (42) にもどろう：

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log P_\theta \{T^{(n)} \geq \theta + \epsilon\} = -\frac{1}{l} D_{\Pi^{(l)}}(\rho_{\theta+\epsilon}^{(l)} \parallel \rho_\theta^{(l)})$$

この式における $\Pi^{(l)}, T^{(n)}$ は、 θ, ϵ, δ に依存して定まる。今、 θ, ϵ, δ を任意に与えて $\Pi^{(l)}, T^{(n)}$ を構成したとし、その後で ϵ だけを動かせば

$$\lim_{\epsilon \downarrow 0} \lim_{n \rightarrow \infty} \frac{1}{\epsilon^2 n} \log P_\theta \{T^{(n)} \geq \theta + \epsilon\} = -\frac{1}{l} \lim_{\epsilon \downarrow 0} \frac{1}{\epsilon^2} D_{\Pi^{(l)}}(\rho_{\theta+\epsilon}^{(l)} \parallel \rho_\theta^{(l)}) \quad (44)$$

が得られる。 $D_{\Pi^{(l)}}(\rho_{\theta+\epsilon}^{(l)} \parallel \rho_{\theta}^{(l)})$ は古典的 Kullback divergence であるから、式 (29) より

$$\lim_{\epsilon \downarrow 0} \frac{1}{\epsilon^2} D_{\Pi^{(l)}}(\rho_{\theta+\epsilon}^{(l)} \parallel \rho_{\theta}^{(l)}) = \frac{1}{2} J_{\theta}^{(l)}[\Pi^{(l)}] \quad (45)$$

となる。ここで右辺は、分布 $\text{Tr} [\rho_{\theta}^{(l)} \Pi^{(l)}(dx)]$ の成す古典的モデル $M^{(l)}[\Pi^{(l)}]$ の Fisher 情報量である。式 (22) (27) より

$$J_{\theta}^{(l)}[\Pi^{(l)}] \leq J_{\theta}^{(l)} = l J_{\theta}$$

となるので、式 (44) (45) より

$$\lim_{\epsilon \downarrow 0} \lim_{n \rightarrow \infty} \frac{1}{\epsilon^2 n} \log P_{\theta} \{T^{(n)} \geq \theta + \epsilon\} \geq -\frac{1}{2} J_{\theta} \quad (46)$$

が成り立つことがわかる。

式 (43) の \inf をとるところで用いた推定量 $T^{(n)}$ は式 (46) を満たすことがわかった。そこで、式 (46) を満たす一致推定量の全体を D_{θ} とおけば、式 (43) の \inf は D_{θ} の要素で達成される：

$$\lim_{\epsilon \downarrow 0} \inf_{\{T^{(n)}\} \in D_{\theta}} \lim_{n \rightarrow \infty} \frac{1}{\epsilon^2 n} \log P_{\theta} \{T^{(n)} \geq \theta + \epsilon\} = -\frac{1}{2} \tilde{J}_{\theta} \quad (47)$$

一方、 $\{T^{(n)}\}$ として式 (27) の \max を達成する測定 Π_{θ} に基づいた MLE をとると、これは式 (46) の等号を達成する。この推定量も D_{θ} の要素であるから、結局

$$\inf_{\{T^{(n)}\} \in D_{\theta}} \lim_{\epsilon \downarrow 0} \lim_{n \rightarrow \infty} \frac{1}{\epsilon^2 n} \log P_{\theta} \{T^{(n)} \geq \theta + \epsilon\} = -\frac{1}{2} \tilde{J}_{\theta} \quad (48)$$

が得られる。

式 (47) と式 (48) とでは $\lim_{\epsilon \downarrow 0}$ と \inf の位置が入れ替わっており、このことから不等式 (32) が導かれる。すなわち、式 (47) (48) は、不等式 (32) の推定論的な意味を示しているとも言える。

残る問題は、 D_{θ} と C とはどの程度異なるのかということである。状態 $\rho_{\theta}^{(n)}$ のもとでの推定量 $T^{(n)}$ の分布の成す古典的モデル $M^{(n)}[T^{(n)}]$ の Fisher 情報量を $J_{\theta}^{(n)}[T^{(n)}]$ とおけば、古典推定理論との類推から、ある種の正則条件のもとでかなり一般的に

$$\lim_{\epsilon \downarrow 0} \lim_{n \rightarrow \infty} \frac{1}{\epsilon^2 n} \log P_{\theta} \{T^{(n)} \geq \theta + \epsilon\} \geq -\frac{1}{2} \lim_{n \rightarrow \infty} \frac{1}{n} J_{\theta}^{(n)}[T^{(n)}] \quad (49)$$

が成り立つと予想される。式 (46) は式 (49) から導かれるので、式 (49) がどのような条件のもとで成り立つのかを明らかにすることが一つの目標になるだろう。

参考文献

- [1] Helstrom, C. W., "Quantum Detection and Estimation Theory," Academic Press, New York, 1976.
- [2] Holevo, A. S., "Probabilistic and Statistical Aspects of Quantum Theory," North-Holland, 1982.
- [3] Yuen, H. P., and Lax, M., "Multiple-parameter Quantum Estimation and Measurement of Nonselfadjoint Observables," Trans. IEEE, IT-19 (1973), 740-750.
- [4] 長岡浩司, "量子統計モデルの Fisher 情報量について," 第10回情報理論とその応用シンポジウム予稿集 (1987), 241-246.
- [5] Nagaoka, H., "An Asymptotically Efficient Estimator for A One-dimensional Parametric Model of Quantum Statistical Operators," Proc. of 1988 IEEE International Symposium on Information Theory (1988), 198.
- [6] 長岡浩司, "Kullback Divergence と Fisher Information の関係について — 古典系から量子系へ — ," 情報理論とその応用学会ジョイントミニワークショップ講演資料 (1992), 63-72.
- [7] Bahadur, R. R., "On the Asymptotic Efficiency of Tests and Estimates," Sankhyā, Vol.22 (1960), 229-252.
- [8] Bucklew, J. A., "Large Deviation Techniques in Decision, Simulation, and Estimation," John Wiley & Sons, 1990.
- [9] 梅垣壽春・大矢雅則, "量子論的エントロピー," 共立出版, 1984.