# 量子推定理論入門

藤原 彰夫

1997 年 6 月 2 日

近年，量子コンピュータが実現すれば大きな整数の素因数分解が現実的な時間で可能となるので，RSA 暗号に代表される現在の暗号システムは安全でないとか，そんな時代が来ても，量子暗号という究極の暗号システムがあるから大丈夫であるとか，量子力学的な原理に基づく情報システムが巷を賑わすようになってきた．本章では，量子情報理論の一分野である量子状態の統計的推定理論をできる限りやさしく紹介してみようと思う．

量子推定理論とは，対象とする物理系について我々が知りうる情報の限界を，統計学の立場から論ずる分野である．とかく華やかな宣伝文句が濫用されがちな量子情報理論の諸分野の中でも，量子推定理論はかなり地味な分野の一つであって，日本語で読める入門的解説も今のところ殆ど出版されていない．しかし，量子力学が元来統計的理論である以上，量子情報の可能性とその限界を正しく把握するためにも，統計学と量子論を融合させた数学的理論体系が必要であることは当然であり，その重要性はもっと広く認知されるべきであろう．このような現状を考慮し，一人でも多くの読者にこの魅力的な分野を紹介したいと思い，この原稿を書いてみた．必要な予備知識は，大学 1 年の時に勉強する線形代数学のごく基本的部分だけである．量子力学については殆ど何も知らないという読者を想定している．

## 1　統計的推定とは

次のような問題を考えてみよう．手元に一枚の変形したコインがある．このコインを投げて表が出る確率を $\theta$，裏が出る確率を $1-\theta$ とするとき，$\theta$ の真の値を知るにはどうすればよいだろうか．実際に何回もコインを投げて，コインを投げた回数で表が出た回数を割れば，$\theta$ に近い値が得られることが予想される．実際，試行回数をどんどん増やせば，真の値 $\theta$ にどんどん近づいていくことは大数の（強）法則の主張するところである．では，この推定方法が一番良い推定方法なのだろうか，それとも他にもっと良い方法があるのだろうか．さらには，そもそも推定方法の良し悪しの基準とはいったい何であろうか．いろいろと疑問が湧いてくる．このような問題を取り扱うのが統計的推定理論であり，そのごく入り口を概観して我々の動機づけを行うのが本節の目的である．

問題をもう少し一般化して，有限事象系 $\Omega = \{\omega_1, \ldots, \omega_r\}$ 上の確率分布 $p_\theta(x)$（$x \in \Omega$）が $d$ 次元パラメータ $\theta = (\theta^1, \ldots, \theta^d) \in \Theta$（$\Theta$ は $\mathbf{R}^d$ のある部分集合）で特徴づけられているとする．そのような確率分布の族

$$\mathcal{S} = \{p_\theta \,;\, \theta \in \Theta\}$$

を，パラメータ $\theta$ を持つ $d$ 次元統計的モデルという．そして，着目する系は，ある $\theta$ の値に対応する確率分布 $p_\theta$ に従ってデータを出力しているのだが，$\theta$ の真の値が未知であるとき，観測データを頼りに $\theta$ の値を推定するのが統計的推定問題である．

$N$ 回の独立試行の結果，データ $(x_1, \ldots, x_N)$（$x_i \in \Omega$）が得られたとする．我々はこれから $\theta$ の値を推定しなければならない．すなわち推定とは，測定データの集合から $\Theta$ への写像

$$\hat{\theta} \,:\, \Omega^N \longrightarrow \Theta$$

を定めることに他ならない．このような写像 $\hat{\theta}$ のことを，パラメータ $\theta$ の推定量という．

さて，推定量は基本的にはどんなものでもよい．例えば前のコインの問題で，どんな測定データが得られても $1/2$ という値を推定値として与える写像も一つの推定量である．しかし，こんな投げやりな推定量では，一般には全然ダメであることは明らかだろう．そこで通常は，推定値の期待値が $\theta$ の真の値に一致するような，性質の良い推定量を考えることが多い．真の分布が $p_\theta$ である時，$N$ 回の独立試行により，データ $(x_1, \ldots, x_N)$ を得る確率を $p_\theta^N(x_1, \ldots, x_N)$ と書けば，

$$p_\theta^N(x_1, \ldots, x_N) = p_\theta(x_1) \cdots p_\theta(x_N) \tag{1}$$

であって，推定量 $\hat{\theta}$ の期待値は

$$E_\theta[\hat{\theta}] := \sum_{(x_1,\ldots,x_N)\in\Omega^N} p_\theta^N(x_1, \ldots, x_N)\, \hat{\theta}(x_1, \ldots, x_N)$$

と書かれる．このとき，全ての $\theta \in \Theta$ に対して $E_\theta[\hat{\theta}] = \theta$ を満たすような推定量 $\hat{\theta}$ のクラスを考えようというわけである．この性質を満たす推定量のことを 不偏推定量 という．

ところで，$p_\theta^N$ 自身を一つのモデルと見なし，$(x_1, \ldots, x_N)$ を一回の測定データと考えることもできるので，$N \to \infty$ とする時の漸近挙動を調べるのでない限り，$N = 1$ と考えても一般性は失われない．そこで，以下では常に $N = 1$ としておく．また，簡単のため，本節では１次元モデルに話を限定する．それは，古典論の場合，１次元モデルと全く同様の議論が，一般の $d$ 次元モデルに対してもそのまま成り立つからである[1]．

さて，推定量の良し悪しの基準には，しばしば平均２乗誤差を用いる．これは

$$\sum_{x\in\Omega} p_\theta(x) \left(\hat{\theta}(x) - \theta\right)^2$$

で定義される量であるが，$\hat{\theta}$ が不偏推定量の場合には，分散

$$V_\theta[\hat{\theta}] := \sum_{x\in\Omega} p_\theta(x) \left(\hat{\theta}(x) - E_\theta[\hat{\theta}]\right)^2$$

[1] 後で分かるように，まさにこの点が古典論と量子論との決定的相違なのである．

と一致する．そして，分散をなるべく小さくするような不偏推定量が良い推定量であるという基準の下で，最良の推定量を捜そうというわけである．この問題に対し，次の Cramér-Rao の定理が基本的である．

**定理 1** (Cramér-Rao)　$p_\theta$ は $\theta$ に関して滑らかであるとする．1 次元モデル $\mathcal{S} = \{p_\theta\}$ のパラメータ $\theta$ に対する任意の不偏推定量を $\hat{\theta}$ とすると，

$$V_\theta[\hat{\theta}] \geq \frac{1}{J_\theta}$$

が成り立つ．ここに $J_\theta$ は

$$J_\theta := \sum_{x \in \Omega} p_\theta(x) \left( \frac{d}{d\theta} \log p_\theta(x) \right)^2$$

で定義される $\theta$ の関数であり，*Fisher* 情報量 とよばれている．ついでながら，$\dfrac{d}{d\theta} \log p_\theta(x)$ は 対数微分 とよばれる．

この定理の量子版をあとで証明するので，ここでは証明を省略する．この定理は，どんなに上手に不偏推定量を選んでも，分散を Fisher 情報量の逆数よりも小さくすることはできないことを意味している．従って，下界 $\dfrac{1}{J_\theta}$ を達成する不偏推定量が存在すれば，それが最も良い推定量ということになる．

はじめのコインの例で具体的に計算してみよう．考える 1 次元モデル $\mathcal{S} = \{p_\theta \, ; \, \theta \in (0,1)\}$ は

$$p_\theta(x) = \begin{cases} \theta, & x = \text{表} \\ 1-\theta, & x = \text{裏} \end{cases}$$

である．これから

$$\frac{d}{d\theta} \log p_\theta(x) = \begin{cases} \dfrac{1}{\theta}, & x = \text{表} \\ -\dfrac{1}{1-\theta}, & x = \text{裏} \end{cases}$$

となるので，Fisher 情報量は

$$J_\theta = \theta \cdot \left( \frac{1}{\theta} \right)^2 + (1-\theta) \cdot \left( -\frac{1}{1-\theta} \right)^2 = \frac{1}{\theta(1-\theta)}$$

と計算され，Cramér-Rao 不等式は

$$V_\theta[\hat{\theta}] \geq \theta(1-\theta)$$

となる．この評価から，$\theta = 1/2$ の時が最も推定しにくく，$\theta$ が 0 または 1 に近づくほど推定しやすいことが分かる．なお，$N$ 回の試行で，表が $r$ 回出たら $r/N$ を推定値とする推定量は，$N = 1$ の場合には，1 回の試行で表が出たら 1，裏が出たら 0 を推定値とする推定量に帰着される．この推定量 $\hat{\theta}$ の期待値は

$$E_\theta[\hat{\theta}] = \theta \cdot 1 + (1-\theta) \cdot 0 = \theta$$

だから不偏推定量であり，

$$V_\theta[\hat{\theta}] = \theta \cdot (1-\theta)^2 + (1-\theta) \cdot (0-\theta)^2 = \theta(1-\theta)$$

だから，Cramér-Rao の下界を達成するという意味で最良の推定量となっている[2]．

以下では，このような統計的推定問題が，量子力学という統計理論の枠組みで，どのように定式化され，古典論とどのように異なるのかを調べていこうと思う．

## 2　量子力学速成コース

ある対象から情報を得るためには，我々はその対象に何らかの働きかけをしなければならない．この働きかけのことを，物理では「観測」とか「測

[2]実際には，1 回だけコインを投げて，パラメータ $\theta$ の値を推定する人などいないであろう．そこで，$N$ 回の独立試行に基づく推定について簡単に補足しておく．この場合には (1) で定義されるモデル $p_\theta^N$ を考えればよい．簡単な計算により，モデル $p_\theta^N$ の Fisher 情報量 $J_\theta^N$ は $J_\theta^N = NJ_\theta$（$J_\theta$ は $p_\theta$ に対する Fisher 情報量）となることが分かり，従って対応する Cramér-Rao 不等式は $V_\theta[\hat{\theta}] \geq 1/NJ_\theta$ となる．例えば，上記コイン投げ実験の場合，Cramér-Rao 不等式は $V_\theta[\hat{\theta}] \geq \theta(1-\theta)/N$ である．そして，$N$ 回の独立試行で表が $r$ 回出たら $r/N$ を推定値とする推定量が，各 $\theta$ で下界 $\theta(1-\theta)/N$ を達成する不偏推定量となっていることも，簡単な計算ですぐ確かめられる．

定」とか呼んでいる．図１は，手元の箱がびっくり箱であるという情報を，Jack が観測によって獲得したときの様子である．

ところで，古典力学的自然描像では，Jack が箱をあける前からその箱がびっくり箱であるという事実（もう少し正確にいえば，誰かが箱を開ければびっくり人形が飛び出すという観測結果）は確定している．すなわち“情報”は観測前に確定しているのであって，本人はただそれを知らないだけ，ということである．ところが，量子力学的自然描像では，Jack の箱の開け方によっては（観測の仕方によっては）宝の山が出てくることもありうる．（観測者によって箱の中味が異なる舌きりすずめの童話は，量子力学を予見していたのかもしれない．）つまり情報は観測前に確定してはおらず，観測してみるまでは分からないというのである．このように，対象となる物理系の「状態」と，そこから我々が情報を引き出す手続きである「測定」とを完全に分離して考えるのが量子力学の特徴である．ただ，このように言ってみても，ただちに反論が出そうである．「サイコロだって振ってみるまでは分からないではないか！」といった具合に．実は，確率論にしても古典論と量子論の違いが厳然とあるのだが，比喩だけで話を進めていては，やはり正しくメッセージが伝わらない．そろそろ数学の出番が来たようである．

量子力学によれば，対象とする物理系に対し，内積を持つ一般には無限次元の複素線形空間 $\mathcal{H}$（複素 Hilbert 空間とよばれる）があって，状態とか測定といった物理的概念は，$\mathcal{H}$ 上の線形作用素（線形写像）によって数学的に表現される．ここでは線形代数の範囲で議論を進めるために，$\mathcal{H}$ が有限次元の場合，すなわち $n$ 次元複素 Euclid 空間 $\mathcal{H}$ で記述される物理系に話題を限定しよう[3]．そして $\mathcal{H}$ の元を $\psi, \varphi$ などのギリシャ文字で書き，$\psi, \varphi \in \mathcal{H}$ の内積を $(\psi, \varphi)$ で表すことにする．なお，ここでは物理の習慣に従い，$\alpha, \beta \in \mathbf{C}$ に対し $(\alpha\psi, \beta\varphi) = \overline{\alpha}\beta(\psi, \varphi)$ としておく．さて，この内

[3]これは一般性を大いに損なう仮定ではあるが，例えばスピン系などの重要な量子系が実際に有限次元 Hilbert 空間上で記述されるという事実など考えると，あながち無茶な仮定というわけでもないのである．

積に関する $\mathcal{H}$ の正規直交基底を任意に一つ固定すれば，$\mathcal{H}$ は $\mathbf{C}^n$ と同一視でき，$\mathcal{H}$ 上の線形作用素は $n$ 次正方行列とみなせることになる．

$n$ 次正方行列の全体の集合を $\mathcal{L}$ と書くことにしよう．$A \in \mathcal{L}$ が与えられたとき，すべての $\psi, \varphi \in \mathcal{H}$ に対して $(\psi, A\varphi) = (A^*\psi, \varphi)$ を満たす行列 $A^* \in \mathcal{L}$ が唯一つ存在する．この $A^*$ は $A$ の随伴行列とよばれ，具体的には $A$ の共役転置行列で与えられる．$A = A^*$ を満たす行列 $A$ のことを Hermite 行列という．任意の Hermite 行列 $A$ が与えられたとし，その固有値 $a_i$（実数）に対応する $\mathcal{H}$ の部分空間への正射影行列を $P_i$ とすれば，$A$ は

$$A = \sum_i a_i P_i$$

とスペクトル分解できる．以下では Hermite 行列が重要な役割を果たすので，その全体の集合を $\mathcal{L}_h$ という記号で表すことにする．

さて，$\mathcal{H}(\cong \mathbf{C}^n)$ という表現の土俵が定まったので，いよいよ状態および測定という概念を導入しよう．天下りではあるが，物理系の 状態 は，次の２条件を満たす $n$ 次正方行列 $S$ で表現される．

(S-1)　$S \in \mathcal{L}_h$　かつ　$S \geq 0$.

(S-2)　$\mathrm{Tr}\, S = 1$.

ここで $S \geq 0$ とは，行列 $S$ が非負定値であること，すなわちすべての $\psi \in \mathcal{H}$ に対して $(\psi, S\psi) \geq 0$ であることを意味する．全ての固有値が非負であると言ってもよい．$S > 0$ も同様に定義される．また，$\mathrm{Tr}\, S$ は行列 $S$ の対角和，すなわち $S = [s_{ij}]$ に対し $\mathrm{Tr}\, S = \sum_{i=1}^{n} s_{ii}$ を表す．なお，任意の $\alpha \in \mathbf{C}$ および任意の $A, B \in \mathcal{L}$ に対し，

$$\mathrm{Tr}\,(\alpha A) = \alpha \mathrm{Tr}\, A, \qquad \mathrm{Tr}\,(A + B) = \mathrm{Tr}\, A + \mathrm{Tr}\, B,$$

$$\mathrm{Tr}\, A^* = \overline{\mathrm{Tr}\, A}, \qquad \mathrm{Tr}\, AB = \mathrm{Tr}\, BA$$

が成り立つ．これらの性質は，以下でしばしば用いられる．(S-1) (S-2) を満たす行列 $S$ は 密度行列 と呼ばれる．状態の数学的表現が密度行列であるといってもよい．

次に，測定という概念について考えよう．ここでは測定結果の集合（事象系）が有限集合の場合についてのみ説明することにする[4]．有限集合 $\Omega = \{\omega_1, \ldots, \omega_r\}$ を測定結果の集合とする測定は，次の２条件を満たす $n$ 次正方行列の集合 $M = \{M_1, \ldots, M_r\}$ で表現される．

(M-1)　各 $i = 1, \ldots, r$ に対し，$M_i \in \mathcal{L}_h$　かつ　$M_i \geq 0$．

(M-2)　$\sum_{i=1}^{r} M_i = I$　（$I$ は単位行列）．

条件 (M-2) の形から，測定 $M$ のことを $I$ の分解ともいう．測定の数学的表現が $I$ の分解であるといってもよい．

さて，状態と測定という物理概念の数学的表現をこうして天下りに導入したわけだが，これら数学的表現から得られる帰結を物理現象に翻訳する手続きがなければ物理学とはいえない．その手続きの一つが量子力学の統計仮説であり，次のように述べられる．

**仮説**　物理系が状態 $S$ にあるとき，測定 $M$ を行った結果として事象 $\omega_i \in \Omega$ が得られる確率 $P(\omega_i)$ は

$$P(\omega_i) = \mathrm{Tr}\, SM_i$$

で与えられる．

我々は，現実世界の状態や測定を，何らかの物理的根拠にもとづいて上で述べた数学的表現に対応づけ，統計仮説に従って実験データを解析することによって，状態や測定の数学的表現が妥当なものであったか否かを検討するのである．現在までのところ，あらゆる観測事実はこの予言通りとなっており，この意味で量子力学は，物理学の基本法則の一つと見なされているのである．

いくつか補足をしておこう．まず，状態および測定に課せられた条件は，$\mathrm{Tr}\, SM_i$ が確率分布を与えるための要請である．実際，状態と測定は互いに独立に勝手に選べるので，$S$ や $M_i$ が非負定値でないと，$\mathrm{Tr}\, SM_i$ が非負

[4] 有限次元 Hilbert 空間上での量子推定問題を扱う場合，有限事象系のみを考えれば十分であることが証明できるので，これは一般性をさらに失う仮定ではない．

の実数にならないことがあり不都合である．また (S-2) (M-2) は確率分布の規格化条件を保証するものである．すなわち

$$\sum_{i=1}^{r} P(\omega_i) = \mathrm{Tr}\, S \left( \sum_{i=1}^{r} M_i \right) = \mathrm{Tr}\, SI = \mathrm{Tr}\, S = 1$$

次に，単純測定という概念を導入しておこう．測定 $M = \{M_1, \ldots, M_r\}$ が 単純測定 であるとは，上記要請 (M-1) (M-2) の他に，さらに次の性質を満たすような測定のこととする．

(M-3)　$M_i^2 = M_i$，かつ $i \neq j$ ならば $M_i M_j = 0$.

これは，$M_1, \ldots, M_r$ が互いに直交する正射影行列であることを意味する．このような測定は von Neumann の意味での測定とも呼ばれ，従来の量子力学の教科書の多くは，このタイプの測定しか扱っていない．これを別の観点から見るために，測定結果の期待値について述べておこう．各事象 $\omega_i$ が実数 $a_i$ に対応づけられているとき，測定結果の期待値 $E[M]$ は

$$E[M] = \sum_{i=1}^{r} a_i P(\omega_i) = \mathrm{Tr}\, S \left( \sum_{i=1}^{r} a_i M_i \right) = \mathrm{Tr}\, SA$$

で与えられる．ここで $A$ は

$$A = \sum_{i=1}^{r} a_i M_i \tag{2}$$

で定義される Hermite 行列である．ここでもし測定 $M$ が単純測定なら，(2) は Hermite 行列 $A$ のスペクトル分解に他ならない．この場合，測定結果の $k$ 次モーメントは

$$\sum_{i=1}^{r} (a_i)^k P(\omega_i) = \mathrm{Tr}\, S \left( \sum_{i=1}^{r} (a_i)^k M_i \right) = \mathrm{Tr}\, S \left( \sum_{i=1}^{r} a_i M_i \right)^k = \mathrm{Tr}\, SA^k \tag{3}$$

と簡単に書くことができる．多くの物理の教科書では，Hermite 行列を物理量（オブザーバブルともいう）とみなし，その固有値を測定データと考える立場をとっているが，これは測定を単純測定に限ることに他ならないのである．この立場では，同時スペクトル分解をもたない（すなわち可換でない）物理量は，同時に測定することができないことになる[5]．しかし，

[5] ここでいう「同時に測定できない」とは，同じ時刻に測定できないという意味ではなく，同時確率分布を定められないという意味である．

物理系から情報を引き出すあらゆる手続きを測定と見なすならば，単純測定では記述できないような測定もたくさん存在する．例えば，精度はある程度犠牲にして，粒子の座標と運動量を同時に測定することは，しばしば行われていることであるが，これも単純測定とはならない．このような一般化された測定は，(M-1) (M-2) のみで特徴づけられることが知られている．そして，互いに非可換な物理量に対応するパラメータを持つモデルに対するパラメータ推定問題は，まさにこのような一般化された測定のクラスで考えることが必要となるのである．なお，任意の一般化測定は，実はある拡張された Hilbert 空間上の単純測定として実現できることも知られており[6]，この事実は一般化測定の物理的実現方法を考える上で重要である．

以上，量子論のエッセンスとしての状態・測定・統計仮説について駆け足で説明してきた．この他にも，測定後の状態に関する射影仮説や状態の時間発展など，重要な話題がまだいろいろあるが，以下の議論には登場しないので，これらは思いきって割愛することにする．

## 3 例

前節の理解を深めるため，簡単な例を２つ述べようと思う．はじめに，サイコロを振るような普通の（古典的）確率論を前節の枠組みで記述してみよう．これにより，古典確率論の特徴が見えてくるはずである．事象 $\omega_i$ が確率 $p_i$ で生起する（$i = 1, \ldots, r$）という状況を考えると，$p_i \geq 0$ および $\sum_{i=1}^{r} p_i = 1$ が成り立ち，$r$ 次元数ベクトル $\boldsymbol{p} = (p_1, \ldots, p_r)$ は $\Omega = \{\omega_1, \ldots, \omega_r\}$ 上の確率分布を与える．この状況を前節の枠組みで記述するには，次のようにすればよい．まず表現空間は $\mathcal{H} \cong \mathbf{C}^r$ とし，状態 $S$ および単純測定 $M = \{M_r\}_{1 \leq i \leq r}$ を次のように定める．

$$
\begin{aligned}
S &= \mathrm{diag}\,[p_1, \ldots, p_r] \\
M_i &= \mathrm{diag}\,[\underbrace{0, \ldots, 0, 1}_{i}, 0, \ldots, 0]
\end{aligned}
$$

[6] Naimark の拡張定理とよばれるものである．

ここに $\mathrm{diag}\,[a_1,\ldots,a_r]$ は，対角要素が順に $a_1,\ldots,a_r$ である対角行列を表すとする．すると統計仮説から，状態 $S$ にある系に対し，測定 $M$ を施した結果，事象 $\omega_i$ を得る確率は

$$P(\omega_i) = \mathrm{Tr}\, SM_i = p_i$$

となり，確かに我々が出発点とした古典的確率分布を与えている．実は古典確率論は，線形作用素の行列表現を任意に固定したとき，全ての状態や測定が対角行列で表される場合に対応していて，状態の対角要素が確率分布を与えるのである．なお，この場合の測定とは，状態の対角要素の１つ１つを取り出してくるための補助的道具に過ぎないから，わざわざ測定などという概念を用いる必要は全くなく，状態のみで記述は完結している．2 節のはじめで「情報は観測前に確定している」と比喩的に表現したのは，こういう意味だったのである．一方，量子力学に従う物理系から情報を引き出す際には，必ずしも状態 $S$ と同時に対角化される（すなわち可換な）測定をおこなう必要はなく，全く独立に選ぶことができるのである．この自由度が量子論の特徴であり，また数学的困難の原因でもある．

次に，最も簡単な量子系として $\mathcal{H} \cong \mathbf{C}^2$ で記述される２準位量子系を考えよう．スピン 1/2 の粒子や光の偏光などが代表的な例である．さて，この系を記述する際には，次の３つの Hermite 行列を用意しておくと便利である．

$$\sigma_x = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \qquad \sigma_y = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}, \qquad \sigma_z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$

これらは Pauli のスピン行列とよばれ，これらと単位行列 $I$ とで，２次正方行列全体を張ることができる．特に，任意の状態 $S$ は次のような線形結合で表される．

$$S = \frac{1}{2}(I + x\sigma_x + y\sigma_y + z\sigma_z) = \frac{1}{2}\begin{bmatrix} 1+z & x-iy \\ x+iy & 1-z \end{bmatrix} \tag{4}$$

まず，$I$ の係数が $1/2$ であるのは条件 (S-2) による．次に，$(x,y,z)$ は Stokes パラメータとよばれる実数の組で，(4) の固有値が

$$\frac{1\pm\sqrt{x^2+y^2+z^2}}{2}$$

であることと条件 (S-1) より

$$x^2+y^2+z^2\leq 1$$

を満たすようにとられる．

次に測定であるが，ここでは代表的な例を紹介するに留める．まず，最も有名な測定として，Stern-Gerlach 型の測定がある．これは $\Omega=\{\pm 1\}$ 上に値をとる測定 $M=\{M(+1),M(-1)\}$ であり，$\alpha^2+\beta^2+\gamma^2=1$ を満たす実数の組 $(\alpha,\beta,\gamma)$ を用いて

$$M(+1)=\frac{1}{2}(I+\alpha\sigma_x+\beta\sigma_y+\gamma\sigma_z)=\frac{1}{2}\begin{bmatrix}1+\gamma & \alpha-i\beta\\ \alpha+i\beta & 1-\gamma\end{bmatrix}$$

$$M(-1)=\frac{1}{2}(I-\alpha\sigma_x-\beta\sigma_y-\gamma\sigma_z)=\frac{1}{2}\begin{bmatrix}1-\gamma & -\alpha+i\beta\\ -\alpha-i\beta & 1+\gamma\end{bmatrix}$$

と定義されるものである．$(\alpha,\beta,\gamma)$ は３次元実 Euclid 空間の方向を表すパラメータであり，このベクトルが表す方向を軸とする（非一様な）磁場の中をスピン $1/2$ の粒子を通過させる実験などに対応する（図２）．統計仮説によれば，測定結果 $\pm 1$ が得られる確率はそれぞれ

$$P(\pm 1)=\frac{1\pm(\alpha x+\beta y+\gamma z)}{2}$$

で与えられる．なお，$M(\pm 1)$ は共に正射影行列であるので，Stern-Gerlach 測定は単純測定である．さらに $\dim\mathcal{H}=2$ の場合，自明でない単純測定は Stern-Gerlach 測定に限られることも容易に分かる．

一般化測定の例としては，例えば次のようなものがある．上記 $(\alpha,\beta,\gamma)$ 方向の Stern-Gerlach 測定 $M=\{M(+1),M(-1)\}$ と，別の方向 $(\alpha',\beta',\gamma')$ に対応する Stern-Gerlach 測定 $M'=\{M'(+1),M'(-1)\}$ の２つの実験装

置を用意しておき，実験者は表の出る確率が $\lambda$，裏の出る確率が $1-\lambda$ のコインを投げて，表が出たら測定 $M$，裏が出たら測定 $M'$ を行うものとする．このような測定は，

$$\Omega = \{(\text{表}, +1),\ (\text{表}, -1),\ (\text{裏}, +1),\ (\text{裏}, -1)\}$$

という4つの事象の上の測定と考えることができ，対応する $I$ の分解は

$$M^{(\lambda)} = \{\lambda M(+1),\ \lambda M(-1),\ (1-\lambda)M'(+1),\ (1-\lambda)M'(-1)\}$$

となる．この $I$ の分解は (M-1) (M-2) の両方を満たすが，$\lambda$ が 0 または 1 でない限り $M^{(\lambda)}$ は正射影からは構成されていないので (M-3) は満たさず，確かに単純測定でない一般化測定を与えている．なお，初めに用意する Stern-Gerlach 測定装置を一般の $k$ 個に増やし，何番目の測定装置を使うかを確率的に決定するような測定を考えることもできる．このように，確率的に測定を混合して作った測定のことを ランダム測定 という．もちろん，ランダム測定でない一般化測定もたくさん存在する．

## 4　1次元パラメータに対する量子推定理論

いよいよ量子推定理論に話題を進めよう．ここではまず，1次元パラメータの推定問題を考えることにし，2次元以上のパラメータ推定問題は次節で扱うことにする．なお，数学的簡単化のため，扱う状態はすべて正定値の密度行列で表現されるものと仮定する[7]．

さて，正定値の密度行列からなる1次元モデル

$$\mathcal{S} = \{S_\theta\ ;\ \theta \in \Theta \subset \mathbf{R}\}$$

が与えられたとし，真の状態はモデル $\{S_\theta\}$ の中にあるが $\theta$ は未知である，という状況の下で $\theta$ を推定する問題を考える．ここではさらに，$S_\theta$ は $\theta$

[7]実は古典的統計学でもしばしば同様の仮定を用いる．すなわち，$\log p_\theta(x)$ を考えるために $p_\theta(x) > 0$ を仮定するのである．$S_\theta > 0$ はこれに対応する条件である．なお，量子論特有の事情により，この条件をはずせる量子統計的モデルも存在し，それはそれで推定理論的に重要なクラスを構成している．

に関して滑らかであると仮定しておく[8]．推定は何らかの測定に基づいてなされるから，パラメータ $\theta$ の推定量は，測定 $M$ と，測定結果から $\Theta$ への写像 $\hat{\theta}$ のペア $(M,\hat{\theta})$ で与えられるとするのが一般的であるが（次節参照），1次元パラメータの推定問題では，結果的に単純測定のみを考えれば十分であることが後で分かるので，ここでは初めからオブザーバブルを推定量とする定式化を行うことにする．すなわち，推定量はある Hermite 行列 $T \in \mathcal{L}_h$ で表される．そして，推定量 $T$ の期待値がすべての $\theta \in \Theta$ に対して真の値 $\theta$ を与える時，すなわち

$$E_\theta[T] := \mathrm{Tr}\, S_\theta T = \theta \tag{5}$$

を満たす時，$T$ を 不偏推定量 という．これを $\theta$ で微分すると

$$\mathrm{Tr}\, \frac{dS_\theta}{d\theta} T = 1 \tag{6}$$

となる．さて，ある固定された $\theta_0 \in \Theta$ で (5) (6) の両方を満たす時，$T$ は $\theta_0$- 局所不偏 であるという．明らかに，全ての $\theta \in \Theta$ に対し $\theta$- 局所不偏である時かつその時に限り，$T$ は不偏である．条件は弱いほうが適用範囲が広いので，以下では専ら $\theta$- 局所不偏性の方を用いる．$\theta$- 局所不偏な推定量 $T$ の分散は，(3)(5) から

$$V_\theta[T] := \mathrm{Tr}\, S_\theta T^2 - (\mathrm{Tr}\, S_\theta T)^2 = \mathrm{Tr}\, S_\theta (T - \theta I)^2$$

と書けることが分かる．

次に，古典推定理論における対数微分の量子版として，各 $\theta \in \Theta$ に対し

$$\frac{dS_\theta}{d\theta} = \frac{1}{2}[S_\theta L_\theta + L_\theta^* S_\theta] \tag{7}$$

を満たす $L_\theta \in \mathcal{L}$ を，$\theta$ における対数微分と見なそう[9]．実際，$S_\theta$ と $L_\theta$ が行列ではなく実数であったならば，(7) の解は

$$L_\theta = \frac{1}{S_\theta}\frac{dS_\theta}{d\theta} = \frac{d}{d\theta}\log S_\theta$$

[8] 行列 $S_\theta$ の微分は，成分ごとの微分で定義する．

[9] $(d/d\theta)\log S_\theta$ の方が，対数微分の自然な量子版を与えるのではないかと考える読者もいることだろうが，この量は今のところ量子推定理論において全く役に立っていない．

となるので，確かに (7) は対数微分の一つの拡張を与えている．なお，(7) の第 1 式の右辺は自動的に Hermite 行列を与える形となっている．さらに，方程式 (7) の解 $L_\theta$ は一意には定まらないことに注意．(7) の解全体の集合を $\mathcal{M}_\theta$ とすれば，以下が成り立つ．

(a)　$\lambda \in \mathbf{R}$，$L_1, L_2 \in \mathcal{M}_\theta$ に対し，$\lambda L_1 + (1-\lambda)L_2 \in \mathcal{M}_\theta$.

(b)　$\mathcal{M}_\theta$ の元のうち，Hermite 行列であるものが唯一つ存在する．

(a) は簡単にチェックできる．また (b) は次のようにして分かる．$\mathcal{L}_h$ から $\mathcal{L}_h$ への実線形写像 $\mathcal{T}$ を

$$\mathcal{T}:\ L \longmapsto \frac{1}{2}[S_\theta L + L S_\theta]$$

で定める．$S_\theta$ は正定値と仮定しているから，$n$ 個の固有値 $\{\lambda_i\}_{1\le i\le n}$（重複も許す）は全て正である．固有値 $\lambda_i$ に対応する固有ベクトル $\psi_i$ からなる $\mathcal{H}$ の一つの基底を $\{\psi_i\}_{1\le i\le n}$ としよう．$K \in \mathrm{Ker}\,\mathcal{T}$ とすると，

$$0 = (\psi_i, (S_\theta K + K S_\theta)\psi_j) = (S_\theta\psi_i, K\psi_j) + (\psi_i, K S_\theta \psi_j) = (\lambda_i + \lambda_j)(\psi_i, K\psi_j)$$

従って全ての $i, j$ で $(\psi_i, K\psi_j) = 0$ となるから $K = 0$．これは $\mathrm{Ker}\,\mathcal{T} = \{0\}$，すなわち $\mathcal{T} : \mathcal{L}_h \to \mathcal{L}_h$ が同型であることを意味している．従って，任意に与えられた Hermite 行列 $dS_\theta/d\theta$ に対し，(7) を満たす Hermite 行列が唯一つ存在する．

(b) のように，(7) の解として一意に定まる Hermite 行列を，$\theta$ における 対称対数微分 とよび $L_\theta^S$ と書く．以下で見るように，対称対数微分は 1 次元パラメータの推定理論において最も重要な概念の 1 つである．

さて，上で定義した対数微分を用いて，Fisher 情報量の量子版を考えよう．そのために，複素線形空間 $\mathcal{L}$ 上に，状態 $S_\theta$ に依存した次の内積を導入する．

$$(A, B)_{S_\theta} := \mathrm{Tr}\, S_\theta B A^*, \qquad A, B \in \mathcal{L}$$

$S_\theta > 0$ と仮定しているので，$(A, A)_{S_\theta} = 0 \Leftrightarrow A = 0$ となり，確かに内積となっている（$\mathcal{H}$ の内積 $(\cdot,\cdot)$ と混乱しないように！）．さて

$$J(L_\theta) := (L_\theta, L_\theta)_{S_\theta}, \qquad L_\theta \in \mathcal{M}_\theta \tag{8}$$

を 量子 Fisher 情報量 とよぼう．これは対数微分のとり方によって値が変わるが，次の点が重要である．

**補題 2**　量子 *Fisher* 情報量 (8) は，$L_\theta = L_\theta^S$ の時かつその時に限り最小となる．

**証明**　$K_\theta = L_\theta - L_\theta^S$ とおけば，$S_\theta K_\theta + K_\theta^* S_\theta = 0$ が成り立つから，

$$\begin{aligned}(L_\theta, L_\theta)_{S_\theta} &= \mathrm{Tr}\, S_\theta (L_\theta^S + K_\theta)(L_\theta^S + K_\theta)^* \\ &= \mathrm{Tr}\, S_\theta L_\theta^S L_\theta^S + \mathrm{Tr}\,(S_\theta K_\theta + K_\theta^* S_\theta) L_\theta^S + \mathrm{Tr}\, S_\theta K_\theta K_\theta^* \\ &= (L_\theta^S, L_\theta^S)_{S_\theta} + (K_\theta, K_\theta)_{S_\theta} \\ &\geq (L_\theta^S, L_\theta^S)_{S_\theta}\end{aligned}$$

等号は，$K_\theta = 0$ の時かつその時のみ成立する．　□

以上の準備の下で，次の重要な結果が導かれる．

**定理 3** (量子 Cramér-Rao)　1 次元モデル $\mathcal{S}$ のパラメータ $\theta$ に対する任意の $\theta$-局所不偏推定量 $T$ に対し，次の不等式が成立する．

$$V_\theta[T] \geq \frac{1}{J(L_\theta)} \tag{9}$$

この下界は $L_\theta$ が対称対数微分 $L_\theta^S$ の時かつその時に限り最大となり，その場合の不等式

$$V_\theta[T] \geq \frac{1}{J(L_\theta^S)} \tag{10}$$

の等号を達成する局所不偏推定量が，各 $\theta$ ごとに存在する．すなわち，不等式 (10) は最良の評価を与えている．

**証明**　分散が $V_\theta[T] = (T - \theta I, T - \theta I)_{S_\theta}$ と表せることに注意すれば，(9) は，内積 $(\cdot,\cdot)_{S_\theta}$ に関する Schwarz の不等式を用いて次のように示される．

$$\begin{aligned}(T - \theta I, T - \theta I)_{S_\theta} (L_\theta, L_\theta)_{S_\theta} &\geq |(T - \theta I, L_\theta)_{S_\theta}|^2 \\ &\geq [\mathrm{Re}\,(T - \theta I, L_\theta)_{S_\theta}]^2\end{aligned}$$

$$
\begin{aligned}
&= \left[\mathrm{Tr}\,\frac{1}{2}(S_\theta L_\theta + L_\theta^* S_\theta)(T-\theta I)\right]^2 \\
&= \left[\mathrm{Tr}\,\frac{dS_\theta}{d\theta}(T-\theta I)\right]^2 \\
&= \left[\mathrm{Tr}\,\frac{dS_\theta}{d\theta}T - \theta\frac{d}{d\theta}(\mathrm{Tr}\,S_\theta)\right]^2 = 1
\end{aligned}
$$

ここで，Re は実部を表し，最後の等号では $\theta$- 局所不偏性条件 (6) および規格化条件 $\mathrm{Tr}\,S_\theta = 1$ を用いた．(9) の下界が $L_\theta = L_\theta^S$ の時かつその時に限り最大となることは，補題 2 から分かる．さらに，任意に与えられた $\theta$ に対し

$$
T = \theta I + \frac{L_\theta^S}{J(L_\theta^S)} \tag{11}
$$

が下界 $1/J(L_\theta^S)$ を達成する $\theta$- 局所不偏推定量であることが，簡単な計算で確かめられる．□

さて，各 $\theta$ ごとに最適な局所不偏推定量があるという状況は，必ずしも満足のいくものではない．実際，対象としている系の状態がどの $\theta$ の値に対応するのかが分からないからこそ $\theta$ の値を推定するのだから，真の値 $\theta$ ごとに最適な推定量があるというだけでは何とも心細い．ある万能な不偏推定量 $T$ があって，これ一つで，全ての $\theta \in \Theta$ に対し，分散の下限 (10) を達成しているという状況が最も好ましいわけだが，残念ながらそのような不偏推定量（これを有効推定量という）はごく限られたモデルに対してしか存在しない．実際，(11) を積分すればすぐ分かるように，有効推定量 $T$ を持つモデルは次のような型のものとなる．

$$
S_\theta = e^{\frac{1}{2}[f(\theta)T - g(\theta)]}\ S_{\theta_0}\ e^{\frac{1}{2}[f(\theta)T - g(\theta)]} \tag{12}
$$

(12) は古典推定理論における指数型分布族の量子系への拡張であり，量子指数型分布族と呼ばれる．これ以外のモデルに対しては有効推定量は存在しないので，古典論と同様，漸近論に持ち込まざるを得ないわけだが，この一般論はまだ完成していない．

量子 Cramér-Rao 不等式の例を見るために，系の時間発展を記述する次の形のモデルを考える．

$$S_t = e^{-itH/\hbar} S_0 e^{itH/\hbar}, \qquad S_0 > 0 \tag{13}$$

ここに $H$ は系のハミルトニアンとよばれる Hermite 行列，$\hbar$ は Planck 定数である．時刻 $t$ を推定する問題を考えよう．(13) を微分すれば

$$L_t = \frac{2i}{\hbar}(H - E_t[H]I) \tag{14}$$

が対数微分の 1 つであることが，ただちに分かる．これに対応する量子 Fisher 情報量は

$$J(L_t) = (L_t, L_t)_{S_t} = \frac{4V_t[H]}{\hbar^2}$$

であるから，時刻 $t$ の任意の局所不偏推定量 $T$ に対し

$$V_t[T] \geq \frac{\hbar^2}{4V_t[H]}$$

が成り立つ．これはいわゆる “時間 – エネルギー不確定性関係” に他ならない．しかも (14) は Hermite 行列でない（従って対称対数微分でない）から，定理 3 より，この下界は決して達成できないことが分かる[10]．このように量子推定理論とは，Heisemberg の不確定性関係のような，誤りではないがそれほど精密な情報を与えない一般原理とは異なり，我々が測定を介してどれだけ正確な情報が得られるかという問題に対する，最も精密なアプローチの 1 つを与えるものなのである．

## 5　パラメータ数が複数の場合

本節では，正定値の密度行列からなる $d$ 次元モデル

$$\mathcal{S} = \{S_\theta \,;\, \theta \in \Theta \subset \mathbf{R}^d\}$$

[10] しかし，純粋状態のような退化した状態に対しては，この限りではないことが知られている．

の $d$ 次元パラメータ $\theta = (\theta^1, \ldots, \theta^d)$ を推定する問題を考える．ここで，$S_\theta$ は $\theta$ に関して滑らかであり，$\left\{\partial S_\theta/\partial\theta^i\right\}_{1\le i\le d}$ は一次独立であると仮定しておく．$M = \{M(\omega_1), \ldots, M(\omega_r)\}$ を有限事象系 $\Omega = \{\omega_1, \ldots, \omega_r\}$ 上の測定とし，対応する確率分布を

$$P_\theta^M(x) = \mathrm{Tr}\, S_\theta M(x), \qquad x \in \Omega \tag{15}$$

と書くことにする．さらに，写像

$$\hat{\theta}:\ \Omega \longrightarrow \Theta$$

を用意し，$M$ と $\hat{\theta}$ のペア $(M, \hat{\theta})$ でもって $d$ 次元パラメータ $\theta$ の推定量を構成する．すなわち，測定 $M$ により事象 $x$ が得られた場合，$\theta$ の推定値として $d$ 次元実ベクトル $\hat{\theta}(x)$ が得られたと解釈するわけである．各 $x \in \Omega$ に対し

$$\hat{\theta}(x) = (\hat{\theta}^1(x), \ldots, \hat{\theta}^d(x))$$

と成分で表すことにする．状態 $S_\theta$ における推定量 $(M, \hat{\theta})$ の期待値ベクトル $E_\theta[M, \hat{\theta}]$ を次式で定義する．

$$E_\theta[M, \hat{\theta}] := \sum_{x\in\Omega} \hat{\theta}(x) P_\theta^M(x) = \left( \sum_{x\in\Omega} \hat{\theta}^1(x) P_\theta^M(x), \ldots, \sum_{x\in\Omega} \hat{\theta}^d(x) P_\theta^M(x) \right)$$

すべての $\theta \in \Theta$ に対し，$(M, \hat{\theta})$ が $E_\theta[M, \hat{\theta}] = \theta$，即ち

$$\sum_{x\in\Omega} \hat{\theta}^i(x) P_\theta^M(x) = \theta^i, \qquad (i = 1, \cdots, d) \tag{16}$$

を満たす時，$(M, \hat{\theta})$ は <u>不偏</u> であるという．これを $\theta^j$ で偏微分して

$$\sum_{x\in\Omega} \hat{\theta}^i(x) \frac{\partial}{\partial\theta^j} P_\theta^M(x) = \delta_j^i, \qquad (i, j = 1, \cdots, d) \tag{17}$$

を得る．さて，$(M, \hat{\theta})$ がある固定された $\theta_0 \in \Theta$ で (16)(17) を満たす時，$(M, \hat{\theta})$ は <u>$\theta_0$- 局所不偏</u> であるという．明らかに，すべての $\theta \in \Theta$ に対し $\theta$-局所不偏である時かつその時に限り $(M, \hat{\theta})$ は不偏である．

ここで，$\theta$ および $\theta$- 局所不偏推定量 $(M,\hat{\theta})$ によって定まる Hermite 行列 $X^i$ （$i=1,\ldots,d$）を次式で定義する.

$$X^i := \sum_{x\in\Omega}(\hat{\theta}^i(x)-\theta^i)M(x) \tag{18}$$

すると，推定量 $(M,\hat{\theta})$ の局所不偏性条件 (16)(17) は，(15) と $\mathrm{Tr}\,S_\theta = 1$，およびその微分 $\mathrm{Tr}\,\dfrac{\partial S_\theta}{\partial\theta^j} = 0$ を用いて

$$\mathrm{Tr}\,S_\theta X^i = 0, \qquad (i=1,\cdots,d) \tag{19}$$

$$\mathrm{Tr}\,\frac{\partial S_\theta}{\partial\theta^j}X^i = \delta^i_j, \qquad (i,j=1,\cdots,d) \tag{20}$$

と同値であることが分かる．また $(M,\hat{\theta})$ を $\theta$- 局所不偏推定量とするとき，状態 $S_\theta$ における推定量 $(M,\hat{\theta})$ の分散行列 $V_\theta[M,\hat{\theta}] = [v^{ij}_\theta] \in \mathbf{R}^{n\times n}$ を次式で定義する.

$$v^{ij}_\theta := \sum_{x\in\Omega}(\hat{\theta}^i(x)-\theta^i)(\hat{\theta}^j(x)-\theta^j)P^M_\theta(x) \tag{21}$$

さて，$d$ 次元モデル $\mathcal{S}$ に対する Cramér-Rao 不等式を導くために，ここで補題を 2つ用意しておく.

**補題 4**　$M$ は有限事象系 $\Omega$ 上の測定，$f$ は $\Omega$ 上の複素数値関数とし，行列 $F$ を

$$F = \sum_{x\in\Omega} f(x)M(x)$$

で定義すると，次の不等式が成立する.

$$\sum_{x\in\Omega}|f(x)|^2M(x) \geq FF^*$$

ここで $A,B\in\mathcal{L}_h$ に対し，$A\geq B$ は $A-B\geq 0$ のことと定義する.

**証明**　任意の行列 $A\in\mathcal{L}$ および任意の非負定値行列 $B\in\mathcal{L}_h$ に対して $ABA^*\geq 0$ であるから[11]

$$\sum_x|f(x)|^2M(x) - FF^* = \sum_x\{f(x)I-F\}M(x)\{f(x)I-F\}^* \geq 0$$

[11] 等号成立条件は $AB=0$ である．この事実は，あとで定理 7 の証明で用いる.

□

次に，実線形空間 $\mathcal{L}_h$ 上に，状態 $S_\theta$ に依存した次の実内積を導入する．

$$\langle A, B\rangle_{S_\theta} := \frac{1}{2}\mathrm{Tr}\, S_\theta(AB + BA), \qquad A, B \in \mathcal{L}_h$$

前節では内積 $(\cdot,\cdot)_{S_\theta}$ を用いて Fisher 情報量を定義したが，$A \in \mathcal{L}_h$ に対しては $(A, A)_{S_\theta} = \langle A, A\rangle_{S_\theta}$ が成り立つので，対称対数微分に基づく Cramér-Rao 不等式 (10) は，実は内積 $\langle\cdot,\cdot\rangle_{S_\theta}$ を用いても導出できることを注意しておく．

**補題 5**　$\theta$ および推定量 $(M, \hat{\theta})$ から (18) により定まる *Hermite* 行列の族 $\{X^i\}_{1\le i\le d}$ に対し，次の不等式が成立する．

$$V_\theta[M, \hat{\theta}] \ge [\langle X^i, X^j\rangle_{S_\theta}] \tag{22}$$

もちろん，これも “左辺 − 右辺 $\ge 0$” を意味する．

**証明**　$\boldsymbol{u} = {}^t(u_1, \ldots, u_d) \in \mathbf{R}^{d\times 1}$ を任意にとる．補題 4 で

$$f(x) = \sum_{i=1}^{d} u_i(\hat{\theta}^i(x) - \theta^i)$$

とおけば，$F = \sum_{i=1}^{d} u_i X^i$ であって

$$\sum_{i,j} u_i u_j \sum_x (\hat{\theta}^i(x) - \theta^i)(\hat{\theta}^j(x) - \theta^j)M(x) \ge \sum_{i,j} u_i u_j X^i X^j \tag{23}$$

が分かる．この両辺に $S_\theta$ を作用させて Tr をとれば

$${}^t\boldsymbol{u} V_\theta[M, \hat{\theta}]\boldsymbol{u} \ge {}^t\boldsymbol{u}[\langle X^i, X^j\rangle_{S_\theta}]\boldsymbol{u} \tag{24}$$

を得る．$\boldsymbol{u}$ は任意だから (22) が示された．　□

以上の準備のもとで，多次元版 Cramér-Rao の定理を導こう．ここでは 1 次元の場合の結果を踏まえ，初めから対称対数微分に基づく定式化を

採用することにする．1 次元の場合と同様，次式でパラメータ $\theta^i$ 方向の対称対数微分 $L^S_{\theta,i} \in \mathcal{L}_h$ を定義する．

$$\frac{\partial S_\theta}{\partial \theta^i} = \frac{1}{2}[S_\theta L^S_{\theta,i} + L^S_{\theta,i} S_\theta] \tag{25}$$

この場合も，各 $L^S_{\theta,i}$ は一意的に定まる．また，$\mathrm{Tr}\, S_\theta = 1$ より $\mathrm{Tr}\, \dfrac{\partial S_\theta}{\partial \theta^i} = 0$ であるから，$\mathrm{Tr}\, S_\theta L^S_{\theta,i} = 0$ が成り立つ．そして，$(i,j)$ 成分が $\langle L^S_{\theta,i}, L^S_{\theta,j} \rangle_{S_\theta}$ で与えられる実対称行列

$$J^S_\theta = [\langle L^S_{\theta,i}, L^S_{\theta,j} \rangle_{S_\theta}]$$

を 対称対数微分 Fisher 情報行列 と呼ぶことにする．

**定理 6**（多次元版量子 Cramér-Rao） $d$ 次元モデル $\mathcal{S}$ のパラメータ $\theta$ に対する任意の $\theta$-局所不偏推定量 $(M, \hat{\theta})$ に対し，次の不等式が成り立つ．

$$V_\theta[M, \hat{\theta}] \geq (J^S_\theta)^{-1} \tag{26}$$

**証明** まず，$\{\partial S_\theta / \partial \theta^i\}_{1 \leq i \leq d}$ は一次独立と仮定しているから，$\{L^S_{\theta,i}\}_{1 \leq i \leq d}$ も一次独立となる．従って，内積 $\langle \cdot, \cdot \rangle_{S_\theta}$ に関する Gram 行列 $J^S_\theta$ は正則であるから，右辺の逆行列は確かに存在することを注意しておく．

さて，局所不偏性条件 (20) を用いれば

$$\langle X^i, L^S_{\theta,j} \rangle_{S_\theta} = \mathrm{Tr}\, \frac{1}{2}(S_\theta L^S_{\theta,j} + L^S_{\theta,j} S_\theta) X^i = \mathrm{Tr}\, \frac{\partial S_\theta}{\partial \theta^j} X^i = \delta^i_j$$

であることがわかる．そこで，$\mathbf{R}^{d \times 1}$ の任意の 2 つの元 $\boldsymbol{u} = {}^t(u_1, \ldots, u_d)$，$\boldsymbol{v} = {}^t(v_1, \ldots, v_d)$ に対し

$$X = \sum_{i=1}^d u_i X^i, \qquad L = \sum_{i=1}^d v_i L^S_{\theta,i}$$

とおくと

$$\langle X, L \rangle_{S_\theta} = \sum_{i,j} u_i v_j \langle X^i, L^S_{\theta,j} \rangle_{S_\theta} = \sum_{i,j} u_i v_j \delta^i_j = {}^t\boldsymbol{u}\boldsymbol{v}$$

となるから， Schwarz の不等式より

$$\begin{aligned} {}^t\boldsymbol{u}[\langle X^i, X^j\rangle_{S_\theta}]\boldsymbol{u} \cdot {}^t\boldsymbol{v}[\langle L^S_{\theta,i}, L^S_{\theta,j}\rangle_{S_\theta}]\boldsymbol{v} &= \langle X, X\rangle_{S_\theta} \cdot \langle L, L\rangle_{S_\theta} \\ &\geq (\langle X, L\rangle_{S_\theta})^2 = ({}^t\boldsymbol{u}\boldsymbol{v})^2 \end{aligned}$$

が得られる．ここで特に $\boldsymbol{v} = (J^S_\theta)^{-1}\boldsymbol{u}$ とおけば，ただちに

$${}^t\boldsymbol{u}[\langle X^i, X^j\rangle_{S_\theta}]\boldsymbol{u} \geq {}^t\boldsymbol{u}(J^S_\theta)^{-1}\boldsymbol{u}$$

が分かる．従って補題 5 とあわせて

$${}^t\boldsymbol{u}V_\theta[M, \hat{\theta}]\boldsymbol{u} \geq {}^t\boldsymbol{u}[\langle X^i, X^j\rangle_{S_\theta}]\boldsymbol{u} \geq {}^t\boldsymbol{u}(J^S_\theta)^{-1}\boldsymbol{u}$$

を得る． □

さて，ここで問題となるのは，定理 6 で得られた分散行列の下界 $(J^S_\theta)^{-1}$ は果たして達成可能な下界なのかどうか，すなわち分散行列をそこまで小さくするような推定量 $(M, \hat{\theta})$ が実際に存在するのかどうかという点である．古典論の場合には，パラメータの次元によらず，任意に与えた一点 $\theta$ において，行列版 Cramér-Rao 不等式の下界を達成する $\theta$- 局所不偏推定量が存在する（これは，以下の定理 7 からも分かる）．量子論の場合でも，パラメータが 1 次元であるならば，対称対数微分による Fisher 情報量の逆数は達成可能な下界を与えていた (定理 3)．そして，一般の次元の場合にも，(26) の下界 $(J^S_\theta)^{-1}$ は次の意味では最良の評価を与えている．

(26) は，全ての $\boldsymbol{u} \in \mathbf{R}^{d\times 1}$ に対し

$${}^t\boldsymbol{u}V_\theta[M, \hat{\theta}]\boldsymbol{u} \geq {}^t\boldsymbol{u}(J^S_\theta)^{-1}\boldsymbol{u} \tag{27}$$

が成り立つことを主張しているわけだが，一方で $\boldsymbol{u} \in \mathbf{R}^{d\times 1}$ を任意に固定した場合，1 次元 Cramér-Rao 不等式 (10) の下界を達成する推定量の存在証明と類似の方法により，

$$\inf_{(M,\hat{\theta})} {}^t\boldsymbol{u}V_\theta[M, \hat{\theta}]\boldsymbol{u} = {}^t\boldsymbol{u}(J^S_\theta)^{-1}\boldsymbol{u} \tag{28}$$

を示すことができる．ここに $\inf_{(M,\hat{\theta})}$ は，全ての $\theta$-局所不偏推定量 $(M,\hat{\theta})$ に対する下限を表す．つまり (28) は，任意に $\boldsymbol{u} \in \mathbf{R}^{d\times 1}$ を与えるごとに，(27) の下界 ${}^t\boldsymbol{u}(J_\theta^S)^{-1}\boldsymbol{u}$ にいくらでも近い値を与える（$\boldsymbol{u}$ に依存した）推定量が存在することを主張している．そしてこの意味で，定理 6 の不等式 (26) は最良の評価を与えているのである．

では，(26) において，行列の等式として $V_\theta[M,\hat{\theta}] = (J_\theta^S)^{-1}$ を満たすような推定量は存在するのだろうか．いいかえれば，(27) の等号を達成する $\theta$-局所不偏推定量 $(M,\hat{\theta})$ であって，しかも $\boldsymbol{u} \in \mathbf{R}^{d\times 1}$ にはよらないものは，果たして存在するのだろうか．実は，そのような推定量は，極めて限られた場合にしか存在しないことが次の定理より分かる．

**定理 7**　$V_\theta[M,\hat{\theta}] = (J_\theta^S)^{-1}$ となる $\theta$-局所不偏な推定量 $(M,\hat{\theta})$ が存在するための必要十分条件は，$\theta$ における対称対数微分 $\{L^S_{\theta,i}\}_{1\le i\le d}$ がすべて互いに可換であることである．

**証明**　定理 6 の証明から分かるように，$V_\theta[M,\hat{\theta}] = (J_\theta^S)^{-1}$ は，すべての $\boldsymbol{u} \in \mathbf{R}^{d\times 1}$ に対し

$$ {}^t\boldsymbol{u}V_\theta[M,\hat{\theta}]\boldsymbol{u} = {}^t\boldsymbol{u}[\langle X^i, X^j\rangle_{S_\theta}]\boldsymbol{u} = {}^t\boldsymbol{u}(J_\theta^S)^{-1}\boldsymbol{u} \tag{29} $$

であることと同値である．はじめに必要性を示そう．

まず，(29) の第 1 の等号，すなわち (24) の等号が成立するための条件を調べよう．我々は $S_\theta > 0$ と仮定しているから，不等式 (23) の下での (24) の等号成立は，(23) 自身の等号成立と同等である．さて，補題 4 の等号の成立条件は

$$ \sum_x \{f(x)I - F\}M(x)\{f(x)I - F\}^* = 0 $$

であって，これは全ての $x \in \Omega$ に対し $\{f(x)I - F\}M(x) = 0$，すなわち

$$ FM(x) = f(x)M(x) $$

と同値である．これを (23) の等号成立条件に翻訳すると，全ての $x \in \Omega$ に対し

$$\sum_{i=1}^{d} u_i X^i M(x) = \sum_{i=1}^{d} u_i (\hat{\theta}^i(x) - \theta^i) M(x)$$

となる．さらに $\boldsymbol{u}$ は任意であるから，結局これは，全ての $i = 1, \ldots, d$ および全ての $x \in \Omega$ に対し

$$X^i M(x) = (\hat{\theta}^i(x) - \theta^i) M(x)$$

であることと同値である．この両辺に左から $X^j$ を作用させると

$$X^j X^i M(x) = (\hat{\theta}^i(x) - \theta^i) X^j M(x) = (\hat{\theta}^i(x) - \theta^i)(\hat{\theta}^j(x) - \theta^j) M(x) = X^i X^j M(x)$$

両辺を $\sum_{x \in \Omega}$ すれば $X^j X^i = X^i X^j$，すなわち $X^1, \ldots, X^d$ は互いに可換であることが分かる．

次に，(29) の第２の等号が成立するための条件を調べよう．定理 6 の証明より，この部分は Schwarz の不等式に他ならないから，等号成立条件は，ある実定数 $c \neq 0$ が存在して $X = cL$，すなわち

$$\sum_{i=1}^{d} u_i X^i = c \sum_{i,j=1}^{d} J^{ij} u_j L^S_{\theta,i}$$

となることである（あとで分かるように，実は $c = 1$）．ここに $J^{ij}$ は $(J^S_\theta)^{-1}$ の $(i,j)$ 成分である．$\boldsymbol{u}$ は任意であるから，特に $u_i = \delta_{ik}$ とおいて

$$X^k = c \sum_{i=1}^{d} J^{ik} L^S_{\theta,i}$$

これを逆に解けば

$$L^S_{\theta,i} = \frac{1}{c} \sum_{k=1}^{d} (J^S_\theta)_{ki} X^k$$

が分かる．この関係と $X^1, \ldots, X^d$ が互いに可換という事実から，(29) の２つの等式が成立するならば，$L^S_{\theta,1}, \ldots, L^S_{\theta,d}$ は互いに可換となることが示された．

今度は十分性を示そう．

$$Y^i := \sum_{k=1}^{d} J^{ki} L^S_{\theta,k}$$

とおけば，$L^S_{\theta,1}, \ldots, L^S_{\theta,d}$ が互いに可換であることから，$Y^1, \ldots, Y^d$ の同時スペクトル分解が存在する．それらを

$$Y^i = \sum_{x \in \Omega} \lambda^i(x) M(x), \qquad i = 1 \ldots, d$$

とし，これに対応して

$$\hat{\theta}^i(x) := \theta^i + \lambda^i(x), \qquad i = 1 \ldots, d$$

とおこう．$M = \{M(x); x \in \Omega\}$，$\hat{\theta}(x) = (\hat{\theta}^1(x), \ldots, \hat{\theta}^d(x))$ とすれば，実はペア $(M, \hat{\theta})$ が下界 $(J^S_\theta)^{-1}$ を達成する $\theta$- 局所不偏推定量となっているのである．実際，

$$Y^i = \sum_{x \in \Omega} (\hat{\theta}^i(x) - \theta^i) M(x)$$

だから $Y^i$ は，$(M, \hat{\theta})$ に対して $(18)$ で定まる $X^i$ に対応しており，局所不偏性 $(19)(20)$ は，

$$\mathrm{Tr}\, S_\theta Y^i = \sum_{k=1}^{d} J^{ki} \,\mathrm{Tr}\, S_\theta L^S_{\theta,k} = 0 \qquad (\because \ \mathrm{Tr}\, S_\theta L^S_{\theta,k} = 0)$$

および

$$\mathrm{Tr} \frac{\partial S_\theta}{\partial \theta^j} Y^i = \sum_{k=1}^{d} J^{ki} \,\mathrm{Tr} \frac{1}{2} (L^S_{\theta,j} S_\theta + S_\theta L^S_{\theta,j}) L^S_{\theta,k} = \sum_{k=1}^{d} J^{ki} (J^S_\theta)_{jk} = \delta^i_j$$

より分かる．また，下界達成は次のように示される．まず，$M(x)$ は正射影であるから，$M(x)^2 = M(x)$ および $M(x)M(x') = 0$ $(x \neq x')$ が成り立つ．従って

$$\begin{aligned} &\sum_x (\hat{\theta}^i(x) - \theta^i)(\hat{\theta}^j(x) - \theta^j) M(x) \\ &\qquad = \left\{ \sum_x (\hat{\theta}^i(x) - \theta^i) M(x) \right\} \left\{ \sum_{x'} (\hat{\theta}^j(x') - \theta^j) M(x') \right\} = Y^i Y^j \end{aligned}$$

そこで，$Y^i$ と $Y^j$ が可換であることに注意して，両辺を $\mathrm{Tr}\, S_\theta[\,\cdot\,]$ すれば

$$v_\theta^{ij} = \langle Y^i, Y^j \rangle_{S_\theta} = \sum_{k,\ell=1}^{d} J^{ki} J^{\ell j} \langle L^S_{\theta,k}, L^S_{\theta,\ell} \rangle_{S_\theta} = \sum_{k,\ell=1}^{d} J^{ki} J^{\ell j} (J^S_\theta)_{k\ell} = J^{ij}$$

を得る．なお，最後の等号では $J^S_\theta$ が実対称行列であることを用いた． □

定理 7 から，行列版 Cramér-Rao 不等式の等号を達成する推定量は，それがたった 1 点 $\theta$ という局所においてさえも一般には存在しないことが分かる．これは，各パラメータ $\theta^i$ の最適な推定量が，一般に互いに非可換な物理量に対応するためであり，von Neumann の意味では同時測定できないことが原因である．この事実こそ量子論特有の非可換性に起因する困難であり，1 次元パラメータ推定問題では現れなかった面白い点なのである．

こうして，複数パラメータの量子推定問題では，行列版量子 Cramér-Rao 不等式 (26) の下界を達成する推定量を捜すという戦略は一般には不毛であることが分かった．すなわち “最適な推定量” という概念の定義を変更せざるを得なくなったわけである．

最適性の別の基準としては，分散行列の対角和[12]

$$\mathrm{tr}\, V_\theta[M, \hat{\theta}] = \sum_i v_\theta^{ii} \tag{30}$$

も考えられる．つまり，$\theta$ における推定の精度を $\mathrm{tr}\, V_\theta[M, \hat{\theta}]$ で評価することにし，この値を最小にするような推定量 $(M, \hat{\theta})$ を捜すのである．この問題も一般には非常に難しい問題であり，今のところいくつかの特殊例について解かれているのみである．この基準の下で最適な推定量が得られている具体例として，2 準位量子系における 2 次元モデルのパラメータ推定問題を，結果のみ紹介しておこう．2 次元パラメータ $\theta = (\theta^1, \theta^2)$ によって

$$S_\theta = \frac{1}{2}(I + x(\theta)\sigma_x + y(\theta)\sigma_y + z(\theta)\sigma_z), \qquad x(\theta)^2 + y(\theta)^2 + z(\theta)^2 < 1$$

と表されるモデルに対し，$\theta$- 局所不偏推定量 $(M, \hat{\theta})$ を最適化することにより，

$$\min_{(M,\hat{\theta})} \mathrm{tr}\, V_\theta[M, \hat{\theta}] = \mathrm{tr}\, (J^S_\theta)^{-1} + \frac{1}{\det J^S_\theta} \mathrm{TrAbs}\, S_\theta[L^S_{\theta,1}, L^S_{\theta,2}]$$

[12] $\mathcal{H}$ におけるトレース Tr と区別するために，ここでは tr という記号を用いた．

が得られる．ここで，$\mathrm{TrAbs}\, A$ は行列 $A$ の固有値の絶対値の総和を表し，$[A, B] := AB - BA$ である．この右辺第２項が量子論特有の非可換性に起因する量子ノイズを表している．なお，この最小値を達成する最適な測定は，$\mathrm{Span}_{\mathbf{R}}\{L^S_{\theta,1}, L^S_{\theta,2}\}$ に属するある線形独立な２つの Hermite 行列をオブザーバブルとみなし，各々が定める単純測定をある比率で混合することにより構成されるランダム測定となる（一意ではなく連続無限個ある）．この事実からも，単純測定を超えて一般化測定のクラスを扱うことの重要性が理解できる．

ここで，量子推定理論の歴史について簡単に触れておこう．レーザー光の発明により光通信技術が注目され，従来の通信理論の限界が議論される中，まず１次元量子推定理論が Helstrom によって定式化された．その後，光コヒーレント信号の受信方式の最適化問題において，Yuen と Lax は，(30) を最適化の基準とする定式化を提唱し，（無限次元ではあるが）調和振動子の Gauss 状態モデルに対する座標・運動量の同時推定問題を解いている．これは，光ヘテロダイン検波の理論的基礎づけを与えるものとなった．その後も，Holevo，長岡，松本，林らの努力により，いくつかの問題が解かれたが，一般的な問題解決への見通しは，今のところ全くたっていない．読者の中から，この問題にチャレンジして見事な理論を打ち立てる者が登場することをひそかに期待している．

## 6　あとがき

本章で紹介したのは，量子情報理論の一分野である量子推定理論の，そのまたほんの入り口にすぎないが，量子論特有の非可換性に起因する難しさと，それが故の面白さを少しでも読者に伝えることができたなら，著者としてこれ以上の喜びはない．最後に，この分野をもっと勉強してみようと思われる読者のために，参考文献をいくつか挙げておこうと思う．

まず，線形代数学の良い教科書はたくさんあるが，量子論とのつながりを強く意識させる個性的な本として [1] [2] を挙げておく．量子推定理論を数学的にきちんと論じている単行本は，和書・洋書を通じて今のところ [3]

のみであろう．この本は見かけよりずっと難解であるが，[4] などを参考にしながら短気を起こさずに読み進めれば，大きな収穫が得られることと思う．また [5] は，量子推定理論の創始者自身が著した分かりやすい本であり，応用物理的色彩が濃い．量子推定も含め，光通信理論の様々な側面を幅広く解説した和書として [6] がある．最後に，分散行列の対角和を最適化の基準とした量子推定の論文として，[7] および [8] を挙げておく．それぞれ，上で述べた調和振動子の Gauss 状態モデル，および２準位量子系の２次元モデルの推定理論を展開したものである．

終りに，量子推定理論という魅力的な分野を著者に紹介して下さった，長岡浩司氏に心からの感謝を申し上げたい．同氏は，本原稿に注意深く目を通され，多くの有益な助言を下さったのみならず，同氏とのディスカッションの結果を原稿に含めることを快諾して下さった．そのため，本原稿はかなりユニークな内容に仕上がったと思う．

## 参考文献

[1] 竹内外史：線形代数と量子力学，裳華房，1981．

[2] 笠原皓司：線形代数と固有値問題，現代数学社，1972．

[3] A. S. Holevo： Probabilistic and Statistical Aspects of Quantum Theory， North-Holland， 1982．

[4] M. Reed and B. Simon： Methods of Modern Mathematical Physics, I: Functional Analysis, Academic Press， 1980．

[5] C. W. Helstrom： Quantum Detection and Estimation Theory， Academic Press， 1976．

[6] 広田 修：光通信理論，森北出版，1985．

[7] H. P. Yuen and M. Lax： Multiple-Parameter Quantum Estimation and Measurement of Non-Selfadjoint Observables， IEEE Trans.，

**IT-19**, pp. 740-750 (1973).

[8] 長岡 浩司：エルミート行列の同時対角化のある一般化とその量子推定理論との関係について，日本応用数理学会論文誌，**1**，pp. 305-318 (1991).