適応的実験計画での二種類の Bahadur 型限界 (II)

林 正人　理化学研究所脳科学総合研究センター 脳数理研究チーム
松本 啓史　科学技術振興事業団 今井量子計算機構プロジェクト

統計学における適応的実験計画, 学習理論における能動学習はともにデータの取り方を適応的に選択できる状況での統計的推測を扱っている. 同種の問題は量子力学に従う系で未知状態を (量子相関を用いない条件の下で) 統計的に推測する際にも表れ, 最適な量子測定の選択が問題になる. 量子系の統計的推測の理論は Helstrom により始められ, 1980 年代の後半に長岡が適応的推定法を導入して以来, 漸近理論の整備が急速にすすんでいる.

本研究では適応的実験計画での推定量を厳密に定式化した上で, その誤差の大偏差型の限界, すなわち Bahadur 型の限界を論じる. 推定量の実用性を無視し, 原理的な限界の厳密な導出に焦点を置いた.

本研究のテーマである「適応的実験計画での二種類の Bahadur 型限界」については著者らによる, 先行研究があるが, その文献には若干の不備があり, また議論に不十分なところがある. 本研究では, 基礎的なところから出発してその文献の不備を修正した上で, その後の発展について若干触れた.

通常の (実験を動かさない) 確率分布族のパラメータ推定の場合, 推定量が真値から ϵ 以上ずれる確率の漸近的な減少率が Bahadur の限界で上から押えられ (Bahadur の不等式) , しかも最尤推定量はこの限界を達成することが知られている. この議論の特徴は, 不等式が 各点での推定量の真値への収束を仮定するだけ証明され, その他の正則性条件が全く必要ないことである. その帰結として,Bahadur の不等式は検定論の Stein の補題から導かれ, 比喩的には推定は二つのわずかに異なる確率分布の間の検定と見ることができる.

ところが,Bahadur の議論を適応的な実験計画に拡張すると, 各点での推定量の収束だけを仮定した場合の限界 $B^w(\theta)$ と「漸近的一様性」を課した場合の限界 $B^s(\theta)$ とでは異なり, 一般には $B^s(\theta) \leq B^w(\theta)$ となる.

以下その詳細を述べる. 本研究では実験結果の確率分布は実験 M と実験対象の状態 ρ の関数となる状況を扱う. 未知パラメータ $\theta \in \Theta \subset \mathbb{R}^d$ でパラメトライズされた状態族 $\mathcal{S} \equiv \{\rho_\theta | \theta \in \Theta \subset \mathbb{R}^d\}$ のパラメータ推定を問題にし, その一次漸近論を扱う. 状態 ρ_θ にある実験 (測定)M を施したときのデータ ω の従う確率分布を $\mathrm{P}_\theta^M(\,\mathrm{d}\omega)$, その分布による確率変数 X の期待値を $\mathrm{E}_\theta^M(X)$ または $\mathrm{E}_\theta^M\, X$ と記す. また, 実験 M の集合を $\mathcal{M}$ と記す.

場合によっては, 複数の実験を確率的に混合して用いることが有用である. このような戦略は $\mathcal{M}$ 上の確率分布 μ で記述される. 以下 $\mathcal{M}$ 上の確率分布全体の集合を $P(\mathcal{M})$ で表し, $P(\mathcal{M})$ の要素も実験という. 混乱の恐れのあるときは $\mathcal{M}$ の要素を基本実験,$P(\mathcal{M})$ の要素を混合実験とよんで区別する. 状態 ρ_θ に実験 $\mu \in P(\mathcal{M})$ を施したときのデータ ω の従う確率分布 $\int_{\mathcal{M}} \mathrm{P}_\theta^M(\,\mathrm{d}\omega)\mu(\,\mathrm{d}M)$ を $\mathrm{P}_\theta^\mu(\,\mathrm{d}\omega)$ と記す.

次に漸近理論を扱うために, 実験回数 n の適応的な実験のもとでの未知パラメータ θ の推定量を定式化する. $k-1$ 番目までの測定値 $\vec{\omega}_{k-1} \equiv (\omega_1, \ldots, \omega_{k-1})$ に依存して k 番目の実験 $\mu_k(\vec{\omega}_{k-1})$ を選ぶ. このような実験の系列の決め方を $\mathcal{E}_n = \{\mu_k(\cdot)\}_{k=1}^n$ で, それによって得られるデータの系列 $\vec{\omega}_n$ の確率分布 $\mathrm{P}_\theta^{\mu_1}(\,\mathrm{d}\omega_1)\mathrm{P}_\theta^{\mu_2(\vec{\omega}_1)}(\,\mathrm{d}\omega_2)\cdots\mathrm{P}_\theta^{\mu_n(\vec{\omega}_{n-1})}(\,\mathrm{d}\omega_n)$ を

$\mathrm{P}_\theta^{\mathcal{E}_n}(\,\mathrm{d}\vec{\omega}_n)$ で表す.n 回同じ実験 μ を繰り返す戦略は $\mu \times n$ と書く. $\vec{\omega}_n$ から推定値 $T \in \mathbb{R}^d$ への関数を $T_n(\vec{\omega}_n)$ と表し, n 回の実験に対する推定量を組 $(\mathcal{E}_n, T_n)$ で定義する. そして, 確率分布族 $\{\mathrm{P}_\theta^\mu | \theta \in \Theta\}$ の Fisher 情報行列を $J_{i,j}^\mu(\theta)$ と略記する.

推定量の列 $\mathcal{E} \equiv \{(\mathcal{E}_n, T_n)\}_{n=1}^\infty$ が以下の条件 **(I)** をみたすとき弱一致とよぶことにする.

(I) $\mathrm{P}_\theta^{\mathcal{E}_n}\{T_n \notin U_{\theta,\epsilon}\} \to 0, \quad \forall\theta \in \Theta, \forall\epsilon > 0.$

そして, $\beta(\mathcal{E}, \theta, \epsilon)$, $\alpha(\mathcal{E}, \theta)$ を以下で定義する.

$$\beta(\mathcal{E}, \theta, \epsilon) \equiv \overline{\lim_{n\to\infty}} \frac{-1}{n} \log \mathrm{P}_\theta^{\mathcal{E}_n}\{T_n \notin U_{\theta,\epsilon}\} \quad (1)$$

$$\alpha(\mathcal{E}, \theta) \equiv \overline{\lim_{\epsilon\downarrow 0}} \frac{1}{\epsilon^2}\beta(\mathcal{E}, \theta, \epsilon). \quad (2)$$

さらに本研究では弱一致性 **(I)** に加えて以下の条件を扱う.

(II) (2) の右辺が lim を持ち, かつ θ について局所一様収束する.

(III) (1) の右辺が lim を持つ.

(IV) $\alpha(\mathcal{E}, \theta)$ が θ に関して連続である.

推定量の系列 $\mathcal{E}$ が弱一致性 **(I)** を持つならば以下の不等式 (3) が成立し, 弱一致性 **(I)** に加えて条件 **(II)** ~ **(IV)** を充たすなら以下の不等式 (4) が成立する.

$$\alpha(\mathcal{E}, \theta_1) \leq B^w(\theta_1), \quad (3)$$

$$\alpha(\mathcal{E}, \theta_1) \leq B^s(\theta_1). \quad (4)$$

2 種の Bahadur 型の限界 $B^w(\theta), B^s(\theta)$ はそれぞれ以下で定義した.

$$B^w(\theta) \equiv \inf_{\|\xi\|=1} \sup_{\mu\in P(\mathcal{M})} \frac{1}{2} J_{i,j}^\mu(\theta)\xi^i\xi^j,$$

$$B^s(\theta) \equiv \sup_{\mu\in P(\mathcal{M})} \inf_{\|\xi\|=1} \frac{1}{2} J_{i,j}^\mu(\theta)\xi^i\xi^j.$$

予稿の付録で示した通り, 両者の限界は各々の意味で達成可能である. しかし, 直観的に自然な推定量の限界は $B^s(\theta)$ であると思われる. 第 1 の理由は, $B^s(\theta)$ の導出に必要な **(II)** ~ **(IV)** の自然さである. 第 2 に, $B^s(\theta)$ は確率分布族の Bahadur の限界の最小化として特徴付けられる. 第 3 に,$B^s(\theta)$ を全ての点で一様に達成する推定量は構成できたが, 本研究で構成した $B^w(\theta)$ を達成する推定量は, 一点でのみ良い効率を達成するに過ぎない.

さらに, 全ての点で $B^w(\theta)$ を達成する推定量は存在しないと著者らは予測している. もしそれが正しければ, $B^s(\theta)$ が推定量の効率の本来の限界といえる. ただし, 適応的実験計画でも未知パラメータが 1 次元ならば $B^s(\theta) = B^w(\theta)$ であるから, 未知パラメータが多次元であることがこの議論で本質的である.

2 つの Bahadur 型の限界が一致するための必要十分条件を予稿の節 3 では求めた. 2 つの限界の一致・不一致は座標系に依存するので, 特定の座標の下での一致, 任意の座標の下での一致, 両者が一致する座標が存在のための必要十分条件を全て導いた. 興味深いことに, 線形の入出力関係という単純な例で, 入力の制限を変えるだけで上記のすべてのケースが現れる. パラメータのとり方に依存しない距離を使って大偏差型の評価を行なう場合には, 両限界の一致・不一致は距離の定義に依存する. この流儀の議論での 2 つの限界の一致条件は本研究で導いた定理の自明な書換えで得られる.

確率分布の推定でモデルが非正則な場合や, 量子相関の利用を許した場合の量子推定では, パラメータが 1 次元であっても, ここで述べた意味での推定と検定のギャップが現れる. これらの場合, 本研究で述べた適応的実験計画とは異なった事情が働いていおり, 全く別の議論が必要である.