Fisher 計量を用いた δ 概十分統計量の定式化とその特徴づけ

立命館大学理工学研究科　山口夏穂里

Kaori Yamaguchi
Graduate School of Science and Engineering, Ritsumeikan University

概要

Ay-Jost-Lê-Schwachhöffer は，統計モデルが与えられたとき誘導モデルの Fisher 計量が元の Fisher 計量と一致する場合，そのモデルにおける統計量は十分であると定義した．本研究では，この十分性を定量的に緩和し，δ 概十分統計量という新たな統計量について紹介する．具体的には，パラメータ空間の任意の接ベクトル v について，$\delta^2\mathfrak{g}(v,v) \leq \mathfrak{g}'(v,v)$ が成り立つものを δ 概十分と定義する．ただし，$\mathfrak{g}$ と $\mathfrak{g}'$ はそれぞれ元のモデルと誘導されたモデルの Fisher 計量を表す．さらに，甘利 – 長岡や Ay–Jost–Lê–Schwachhöfer による単調性定理に基づき，こうした統計量の誘導モデルにおける Fisher 計量 $\mathfrak{g}'$ は，元の Fisher 計量 $\mathfrak{g}$ と Lipschitz 同値であるため，この統計量による情報損失は一様に抑えられることが示される．さらに，このような概十分統計量を条件付き確率や密度関数のある種の分解によって特徴づける．

1　導入

十分統計量は，1922 年に統計学者 Fisher [9] によって導入されて以来，統計学や情報理論における基本的な概念として広く知られている．Fisher は，パラメータ付けされた確率分布の族に対して，推定に必要な情報を完全に保持する統計量に注目した [9]．その後，Neyman がこのような統計量を定式化し，十分統計量と名付けた [14]．さらに，甘利と長岡 [1] は Fisher 計量を用いて，十分統計量の特徴づけを行った．[1] によれば，統計量によって誘導されるモデルの Fisher 計量が元のモデルの Fisher 計量と一致することと，その統計量が十分であることは同値である．その後 Ay–Jost–Lê–Schwachhöfer [2, 5] は，多様体が無限次元の場合の統計モデルを厳密に定式化した．この枠組みのもとで密度関数に基づく十分統計量の新たな特徴づけを行い，甘利 – 長岡の結果を一般化した．また，Kullback [11] も関連する成果を得ており，Kullback–Leibler ダイバージェンスを用いて十分統計量を特徴づけている．

本稿では，誘導されたモデルの Fisher 計量が元のモデルの Fisher 計量と双 Lipschitz 同値となるような統計量を δ 概十分統計量として導入した [17] の内容を紹介する．またその主結果として，[5] における結果とパラレルに密度関数の観点から δ 概十分統計量を特徴づける．十分統計量は，Rao–Blackwell–Kolmogorov の定理 [7, 10, 16] や Lehmann–Scheffé の定理 [13] によって，統計モデルの推定量を分散を小さくするという意味で改良するために利用できることから，δ 概十分統計量は，十分統計量が存在しないモデルにおいて有用となり得る．

まず，Ay–Jost–Lê-Schwachhöfer [2, 5] によるパラメータ付き測度モデルと Fisher 計量を用いた十分統計量の定義を簡単に述べる．

定義 1.1 (パラメータ付き測度モデル)**.** パラメータ付き測度モデルとは，次を満たす三つ組 $(M, \Omega, \mathbf{p})$ のことである．

- 有限次元または無限次元 Banach 多様体 M,

- 可測空間 Ω,
- C^1 級写像 $\mathbf{p} : M \to \mathcal{M}(\Omega)$　ただし，$\mathcal{M}(\Omega) := \{\mu : \mu \text{ は } \Omega \text{ 上有限測度}\}$.

パラメータ付き測度モデルとは，一言で述べると，Banach 多様体の点でパラメトライズされた確率密度関数の族である．また，上の定義では Banach 多様体という言葉が出てきているが，以降は有限次元の場合で話を進めるので単なる多様体と思っていただいても構わない.

このパラメータ付き測度モデルの例として，以下のような n 回のコイン投げのモデルを紹介する．以下の例では，コインの表が出る確率 $\xi \in M$ がパラメータである.

例 1.2. 三つ組 $(M, \Omega, \mathbf{p})$ として，以下を考える．$n \in \mathbb{N}$ に対して，

- $M = (0, 1)$,
- $\Omega = \{0, 1\}^n$,
- $\mathbf{p}(\xi) = \xi^{a(\omega)}(1-\xi)^{n-a(\omega)}\mu_0$，$\omega \in \Omega$　(ただし，μ_0 は数え上げ測度，$a(\omega)$ は ω に現れる 1 の数)．

するとこのとき，この三つ組はパラメータ付き測度モデルである.

次に，統計量の定義を紹介する.

定義 1.3 (統計量)**.** 可測空間 Ω と Ω' に対して, 可測写像 $a : \Omega \to \Omega'$ を統計量という.

ここで，$\Omega' = \{0, 1, ..., n\}$ とすると（ Ω' は可測空間)，例 1.2 で与えた写像 a は Ω から Ω' への統計量になっている．$\mathbf{p}'(\xi) = a_*\mathbf{p}(\xi)$ とすると，次のような別のパラメータ付き測度モデルを誘導する.

例 1.4. 三つ組 $(M, \Omega', \mathbf{p}')$ として，

- $M = (0, 1)$,
- $\Omega' = \{0, 1, ..., n\}$,
- $\mathbf{p}'(\xi) = {}_nC_k \xi^k (1-\xi)^{n-k}\mu_0$ $(\forall k \in \Omega')$　(ただし，μ_0 は数え上げ測度).

この三つ組 $(M, \Omega', \mathbf{p}')$ を，例 1.2 の $(M, \Omega, \mathbf{p})$ から，統計量 a によって誘導されたモデルという．2 つのコイン投げのモデルが登場したが，前者はコインの表がいつ出るかを考慮したものであり，後者はコインの表がいつ出るかという情報をそぎ落とし，コインの表が何回出たかについてのみ着目したモデルである.

一般に元のモデル $(M, \Omega, \mathbf{p})$ から統計量 $a : \Omega \to \Omega'$ によって誘導されるモデル $(M, \Omega', \mathbf{p}')$ は多様体 M は変わらず，統計量によって可測空間が取り替えられ，測度の族は $\mathbf{p}' = a_*\mathbf{p}$ によって定義される.

次に，Jost らが十分統計量を情報幾何学の分野で再定義するのに用いた Fisher 計量を紹介するが，それには，測度の根，k 次可積分性という概念が必要なので，測度の根を説明した後に k 次可積分性の定義を紹介する.

Ω 上の有限測度の集合 $\mathcal{M}(\Omega)$ を順序関係 $\leq$ を持つ順序集合として考える．ただし，$\omega_1, \omega_2 \in \mathcal{M}(\Omega)$ に対して，$\omega_1 \leq \omega_2$ とは ω_1 が ω_2 に dominate されているときをいう．$r > 0$ に対して, $\mathcal{M}(\Omega)$ 上の帰納極限

$$\mathcal{S}^r(\Omega) := \lim_{\to} L^{1/r}(\Omega, \mu)$$

を考える．ただし，$L^{1/r}(\Omega, \mu)$ とは，固定された $r \in (0, 1]$ に対して，$\int_\Omega |\phi|^{1/r} d\mu < \infty$ を満たす Ω 上可測関数全体のなす空間である．ここで $\mu_1 \leq \mu_2$ のとき, 包含写像 $L^{1/r}(\Omega, \mu_1) \to L^{1/r}(\Omega, \mu_2)$ は Banach 空間の等長写像になっている．したがって，$\mathcal{S}^r(\Omega)$ は Banach 空間の構造を持つ．$\phi \in L^{1/r}(\Omega, \mu)$ に対して, 写像 $L^{1/r}(\Omega, \mu) \to \mathcal{S}^r(\Omega)$ の像を $\phi\mu^r$ と表し，このようにして μ の r 根 μ^r は定義される．また，$r > 0$ に対して，

$$\mathcal{M}^r(\Omega) := \{\, \phi\mu^r \in \mathcal{S}^r(\Omega) \mid \mu \in \mathcal{M}(\Omega), \phi \geq 0 \,\}$$

とおく．

定義 1.5. $k \in \mathbb{N}$ に対して, パラメータ付き測度モデル $(M, \Omega, \mathbf{p})$ が k 次可積分であるとは，写像

$$\mathbf{p}^{1/k} : M \to \mathcal{M}^{1/k}(\Omega), \qquad \xi \mapsto \mathbf{p}(\xi)^{1/k}$$

が C^1 級であるときにいう．

本稿では，2 次可積分モデルに焦点を当てる．$(M, \Omega, \mathbf{p})$ が 2 次可積分で，$\mathbf{p}(\xi)$ が密度関数 p によって $\mathbf{p}(\xi) = p(\omega; \xi)\mu_0$ という形で表される場合を考える．このとき，M 上の Fisher 計量は以下で定義される．

定義 1.6 (Fisher 計量)**.** $(M, \Omega, \mathbf{p})$ を 2 次可積分パラメータ付き測度モデルとする．また，Ω 上の有限測度 μ_0 と密度関数 p に対して，$\mathbf{p}$ が $\mathbf{p}(\xi) = p(\omega; \xi)\mu_0$ と書けるとする．このとき M 上の Fisher 計量 $\mathfrak{g}$ は M 上の任意の接ベクトル v,w に対して，以下の式で定義される．

$$\mathfrak{g}(v, w) = \int_\Omega (\partial_v \log p(\cdot; \xi))(\partial_w \log p(\cdot; \xi)) d\mathbf{p}(\xi). \tag{1.1}$$

2 次可積分性の仮定により，この積分は適切に定義される [5]．M 上の接ベクトル v に対して，統計量 κ の (2 次) 情報損失は $\mathfrak{g}(v, v) - \mathfrak{g}'(v, v)$ と定義される [1]. ただし, $\mathfrak{g}'$ は κ によって誘導されるモデル $(M, \Omega', \kappa_*\mathbf{p})$ の Fisher 計量である．情報損失は，Fisher 計量に関する単調性定理 [1, 3, 4, 12] により非負となり，

$$\mathfrak{g}'(v, v) \leq \mathfrak{g}(v, v) \tag{1.2}$$

が M 上の任意の接ベクトル v に対して成り立つ．Ay–Jost–Lê–Schwachhöfer によると，統計量 κ が十分であるとは，情報損失がすべての点で 0 になる場合，すなわち M 上の任意の接ベクトル v に対して (1.2) の等号が成立することを意味する．

定義 1.7 (十分統計量)**.** 2 次可積分パラメータ付き測度モデル $(M, \Omega, \mathbf{p})$ 上の統計量 κ が十分であるとは，M 上の任意の接ベクトル v に対して，

$$\mathfrak{g}'(v, v) = \mathfrak{g}(v, v)$$

が成り立つときにいう．

n 回のコイン投げを考えたとき，例 1.2 で定義した a は十分統計量である．ノルムを計算するとそれが確かめられる．

[17] の主結果においては Jost らが定義した十分統計量を定量的に一般化した δ 概十分統計量を導入し，その特徴づけを得た．まず，δ 概十分統計量の定義を紹介する．

2 δ 概十分統計量とその特徴づけ

定義 2.1 (δ 概十分統計量)**.** $(M, \Omega, \mathbf{p})$ を 2 次可積分パラメータ付き測度モデル, Ω' を可測空間とし, $0 < \delta \leq 1$ とする．このとき，統計量 $\kappa : \Omega \to \Omega'$ が $(M, \Omega, \mathbf{p})$ に対して δ 概十分であるとは，M 上のすべての接ベクトル v に対して次が成り立つことを意味する．

$$\delta^2 \mathfrak{g}(v, v) \leq \mathfrak{g}'(v, v). \tag{2.1}$$

ここで，$\mathfrak{g}$ と $\mathfrak{g}'$ はそれぞれ $(M, \Omega, \mathbf{p})$ と κ によって誘導されるモデル $(M, \Omega', \kappa_*\mathbf{p})$ によって定義される M 上の Fisher 計量である．

単調性定理 (1.2) より $\mathfrak{g}'(v,v) \le \mathfrak{g}(v,v)$ であるから,

$$\delta^2\mathfrak{g}(v,v) \le \mathfrak{g}'(v,v) \le \mathfrak{g}(v,v)$$

であり，つまり $\mathfrak{g}'$ は $\mathfrak{g}$ と双 Lipschitz 同値である．δ 概十分統計量は，十分統計量が存在しない場合に有用である．いくつかの十分統計量のクラス，例えばベイズ十分統計量 [6] や線形十分統計量 [15, 8] が研究されていることに留意されたい．

δ 概十分統計量の例を以下に示す．

例 2.2 (二項分布)**.** $n \in \mathbb{N}$ に対して,

- $M = (0,1)$,
- $\Omega = \{0,1\}^n$（数え上げ測度を備える),
- $\mathbf{p}(\xi)(\{\omega\}) = \xi^{a(\omega)}(1-\xi)^{n-a(\omega)} \quad (\xi \in M,\, \omega \in \Omega)$, ただし $a(\omega)$ は ω が含む 1 の数とする.)

$1 \le m \le n$ に対して, $\Omega' = \{0,1\}^m$ とし，また $\kappa\colon \Omega \to \Omega'$ を

$$\kappa(\omega_1,\ldots,\omega_n) = (\omega_1,\ldots,\omega_m)$$

により定める．すると κ は $\sqrt{\frac{m}{n}}$ 概十分である ([17] 参照). 実際，$\frac{m}{n}\mathfrak{g}(v,v) = \mathfrak{g}'(v,v)$ が成り立つ.

以下の例は，浅岡正幸氏によって与えられたものである．

例 2.3. パラメータ付き測度モデルとして，以下を考える．$M = (0,\frac{1}{2})$, $\Omega = \{0,1\}^2$，$\mathbf{p}(\xi) = p(\cdot;\xi)\mu_0$, ただし μ_0 は Ω 上数え上げ測度．また，密度関数 p を以下で与える．$0 < \xi < \frac{1}{2}$ に対して，

$$p((0,0);\xi) = \frac{1}{2} - \xi, \qquad p((0,1);\xi) = \xi,$$
$$p((1,0);\xi) = \xi^2, \qquad p((1,1);\xi) = \frac{1}{2} - \xi^2.$$

これは、2 回目のコイン投げの結果が最初のコイン投げに依存するような 2 枚のコインのトスのモデルとみなすことができる．

一方，$\Omega' = \{0,1,2\}$ とし，統計量 $\kappa : \Omega \to \Omega'$ を $\kappa((x,y)) = x + y$ と定義する．すると，誘導されたモデルの $\kappa_*\mu_0$ に対する密度関数 p' は以下になる．

$$\begin{aligned}
p'(0;\xi) =& p((0,0);\xi) = \frac{1}{2} - \xi,\\
p'(1;\xi) =& \frac{1}{2}\big(p((0,1);\xi) + p((1,0);\xi)\big) = \frac{\xi+\xi^2}{2},\\
p'(2;\xi) =& p((1,1);\xi) = \frac{1}{2} - \xi^2.
\end{aligned}$$

よって,

$$\begin{aligned}
\mathfrak{g}(\partial_\xi,\partial_\xi) = \|\partial_\xi \log p(\cdot;\xi)\|^2 &= \frac{1}{\frac{1}{2}-\xi} + \frac{5}{\xi} + \frac{4\xi^2}{\frac{1}{2}-\xi^2},\\
\mathfrak{g}'(\partial_\xi,\partial_\xi) = \|\partial_\xi \log \kappa^* p'(\cdot;\xi)\|^2 &= \frac{1}{\frac{1}{2}-\xi} + \frac{(2\xi+1)^2}{2\xi(\xi+1)} + \frac{4\xi^2}{\frac{1}{2}-\xi^2}
\end{aligned}$$

となり, $0 < \forall\xi < 1/2$ に対して $\frac{1}{10}\|\partial_\xi \log p(\cdot;\xi)\|^2 < \|\partial_\xi \log \kappa^* p'(\cdot;\xi)\|^2$ となる. さらには, $\frac{\|\partial_\xi \log \kappa^* p'(\cdot;\xi)\|^2}{\|\partial_\xi \log p(\cdot;\xi)\|^2} \to \frac{1}{10}(\xi \to 0)$ がわかる．よって，κ は $\frac{1}{\sqrt{10}}$ 概十分統計量であり, $\delta > \frac{1}{\sqrt{10}}$ のとき δ 概十分統計量でない.

一方で，任意の $0 < \delta \leq 1$ に対して δ 概十分でない統計量については，例 3.5 を参照されたい．

Ay–Jost–Lê-Schwachhöfer [5] は，十分統計量が Fisher–Neyman の十分統計量と密接に関連していることを示すために，条件付き確率の観点から十分統計量を以下のように特徴づけた．

定理 2.4 ([11],[3],[5, Chapter 5], Fisher–Neyman の特徴づけ [14])**.** $(M, \Omega, \mathbf{p})$ を 2 次可積分パラメータ付き測度モデルで p が正なる Ω 上有限測度 μ_0 に対して

$$\mathbf{p}(\xi) = p(\omega;\xi)\mu_0$$

と表せるものとする．また，$\kappa : \Omega \to \Omega'$ を統計量とし，$p'(\omega;\xi)$ を $\kappa_*\mu_0$ に関する $\kappa_*\mathbf{p}(\xi)$ の密度関数とする．さらに，M を連結と仮定する．このとき以下は同値である:

(i) κ が $(M, \Omega, \mathbf{p})$ に対して十分である．

(ii) 任意の M 上接ベクトル v に対して，$\partial_v \log p = \partial_v \log \kappa^* p'$.

(iii) 写像 $\xi \mapsto \frac{p(\cdot;\xi)}{p'(\kappa(\cdot);\xi)}$ が定値写像.

(iv) 関数 $s : \Omega' \times M \to \mathbb{R}$ と関数 $t \in L^1(\Omega, \mu_0)$ が存在して以下を満たす.

$$p(\omega;\xi) = s(\kappa(\omega);\xi)t(\omega), \qquad \mu_0\text{-}a.e.\ \omega \in \Omega,\ \forall \xi \in M.$$

条件 (iv) は，Fisher–Neyman 十分統計量と密接に関連している．統計量 κ が Fisher–Neyman 十分であるとは，Ω 上に有限測度 μ_0 が存在して，ある $q(\cdot;\xi) \in L^1(\Omega', \mu_0')$ に対して

$$\mathbf{p}(\xi) = q(\kappa(\cdot);\xi)\mu_0'$$

が成り立つことを指す．ここで，$\mu_0' = \kappa_*\mu_0$ とする．実際，定義から明らかなように，κ が Fisher–Neyman 十分であることは，条件 (iv) に加えて $s(\cdot;\xi) \in L^1(\Omega', \kappa_*\mu_0)$ がすべての $\xi \in M$ に対して成り立つことと同値である（[5, Theorem 5.3]）．

パラメータ付き測度モデルの利点の一つは，支配測度の存在を必要としないことである（[5, Chapter 3] ）．しかし，定理 2.4 においては，支配測度 μ_0 の存在の仮定が本質的であることが知られている．Ay–Jost–Lê-Schwachhöfer は，支配測度が存在しないパラメータ付き測度モデル上の統計量の例を示し，それが条件 (i) を満たすが，他の条件 (ii)，(iii)，および (iv) のいずれも満たさないことを示した（[5, Example 5.6]）．

我々の主結果 [17] は，定理 2.4 にパラレルな δ 概十分統計量の次の特徴づけである：

定理 2.5 ([17])**.** $(M, \Omega, \mathbf{p})$ を，密度関数 p が正になるような Ω 上の有限測度 μ_0 に対して $\mathbf{p}(\xi) = p(\omega;\xi)\mu_0$ の形で書ける 2 次可積分パラメータ付き測度モデルとする．また，M を Fisher 計量を備える C^2 級有限次元多様体とする．さらに，統計量 $\kappa : \Omega \to \Omega'$ および $0 < \delta \leq 1$ に対して，以下は同値である．

(i) κ が $(M, \Omega, \mathbf{p})$ に対して δ 概十分である．

(ii) M の任意の接ベクトル v に対して，

$$\left\| \partial_v \log \frac{p(\cdot;\xi)}{\kappa^* p'(\cdot;\xi)} \right\| \leq \sqrt{1-\delta^2} \|\partial_v \log p(\cdot;\xi)\| \tag{2.2}$$

が成り立つ．ただし，$\|\cdot\|$ は L^2 ノルムである．

(iii) 写像 $M \to L^2(\Omega, \mu_0); \xi \mapsto \log \frac{p(\cdot;\xi)}{p'(\kappa(\cdot);\xi)}$ が Fisher-Rao 距離，また Fisher 計量と L^2 計量によって定義される距離に対して $\sqrt{1-\delta^2}$ 局所 Lipschitz である．

(iv) $\exists s : \Omega' \times M \to \mathbb{R}, \exists t : \Omega \times M \to \mathbb{R}_{>0}$ s.t. $\log t(\cdot;\xi) \in L^2(\Omega, \mu_0)\ \forall \xi \in M$,

- $p(\omega;\xi) = s(\kappa(\omega);\xi)t(\omega;\xi)$ μ_0-a.e. $\omega \in \Omega, \forall \xi \in M$,

- 写像 $M \to L^2(\Omega, \mu_0); \xi \mapsto \log t(\cdot; \xi)$ が Fisher-Rao 距離，また Fisher 計量と L^2 計量によって定義される距離に対して $\sqrt{1-\delta^2}$ 局所 Lipschitz である．

ただし，Fisher-Rao 距離は Fisher 計量から誘導される距離である．

定理 2.4 は $k=2$ 以外の場合の k 次可積分パラメータ付き測度モデルに対して成り立つ．しかし，定理 2.5 を 2 以外の k に拡張できるかどうかはわかっていない．また，$\delta = 1$ の場合，条件 (iv) は定理 2.4 の条件 (iv) と一致しないことにも注意してほしい．

また $\dim M < \infty$ であり，Fisher 計量が C^2 級であると仮定する必要がある．これより，すべての接ベクトルは測地線に接し，すべての M の点には開近傍であって，その上のすべての点の組は測地線で繋ぐことができるものが存在する．証明が示唆するように，M 上のすべての点のペアが測地線で繋がれている場合，局所 Lipschitz 条件を Lipschitz 条件に置き換えることができる．

定理 2.4 の場合と同様に，支配測度 μ_0 の存在の仮定は定理 2.5 で不可欠である．[5, 例 5.6] の統計量 κ は十分であるが，任意の $\delta > 0$ に対して条件 (iv) を満たさないことがわかる．実際，[5, 例 5.6] のモデルの密度関数が $p(\omega; \xi) = s(\kappa(\omega); \xi) t(\omega; \xi)$ と分解されるとき $\log t$ は $\xi = 0$ で局所 Lipschitz ではない．

3 二項分布の概十分統計量

この章では指数分布族の基本的な例である二項分布の統計モデルに対する概十分統計量の例をいくつか示す．

有限の標本空間を持つ統計モデルの場合，定義により，統計量 κ によって誘導されたモデルの Fisher 計量は，標本空間の分割 $\Omega = \sqcup_{\omega' \in \Omega'} \kappa^{-1}(\omega')$ のみに依存する．したがって，本質的に，このようなモデルには相違なる統計量は有限個しか存在しない．定理 2.4 を使えば，指数分布族のすべての十分統計量を分類するのは容易である．実際，κ が十分であるとき，各 $\omega' \in \Omega'$ に対して，すべての要素 $\omega \in \kappa^{-1}(\omega')$ は，互いにスカラー倍である密度関数 $p(\omega; \cdot)$ を持つ．二項分布の場合，1 の数を数える統計量 $a \colon \Omega \to \{0, 1, \dots, n\}$ が十分であることはよく知られている．密度関数 $\xi^i (1-\xi)^{n-i}$ と $\xi^j (1-\xi)^{n-j}$ $(i \neq j)$ は互いの倍数ではないため，統計量 a は，他の任意の十分統計量が a によって与えられる分割の細分を定義するという意味で，十分統計量の中で最小である．したがって，このモデルには唯一の最小十分統計量 [13] が存在する．

次に，概十分統計量の例を紹介する．二項分布のモデル $(M, \Omega, \mathbf{p})$ を考える．すなわち，支配測度 μ_0 および $\mathbf{p}(\xi)(\omega) = \xi^{a(\omega)} (1-\xi)^{n-a(\omega)}$ $(\xi \in M,\ \omega \in \Omega)$ で $M = (0, 1)$，$\Omega = \{0, 1\}^n$ として与えられる．ここで，$a(\omega)$ は ω の要素に含まれる 1 の数を表す．この場合，概十分統計量の典型的な例は例 2.2 である．$m = 1$ の場合，この統計量は $0, 1$ の値を取るため，最小になる．最小のものに焦点を当てるために，$0, 1$ の値を持つ統計量を考える．これをバイナリー統計量と呼ぶ．

命題 3.1. $\kappa \colon \{0,1\}^n \to \{0,1\}$ を二項分布のモデル上の統計量とする（例 2.2）．$f(\xi) = \kappa_* \mathbf{p}(\xi)(\{0\})$ とする．κ が全射であり，$\partial_\xi f(\xi)$ が $[0, 1]$ 上で零点を持たないときかつそのときに限り，κ は何らかの $\delta > 0$ に対して δ 概十分になる．

命題 3.1 を証明するために，まず $f(\xi)$ に関して Fisher 計量を次のように計算する．

補題 3.2. κ が全射である場合，κ によって誘導されるモデルの Fisher 計量 $\mathfrak{g}'$ は次のように与えられる．

$$\mathfrak{g}'(\partial_\xi, \partial_\xi) = \frac{\partial_\xi f(\xi)^2}{f(\xi)\bigl(1 - f(\xi)\bigr)}.$$

証明. 仮定により，$C_0 = \kappa_* \mu_0(\{0\})$，$C_1 = \kappa_* \mu_0(\{1\})$ が正である．$f(\xi) = \kappa_* \mathbf{p}(\xi)(\{0\}) = C_0 p'(0; \xi)$ によ

り，$p'(0;\xi)=\frac{f(\xi)}{C_0}$ となる．$\kappa_*\mathbf{p}(\xi)$ は Ω' 上の確率測度であるため，$p'(1;\xi)=\frac{1-f(\xi)}{C_1}$ となり，したがって

$$\begin{aligned}\mathfrak{g}'(\partial_\xi,\partial_\xi)&=\|\partial_\xi\log\kappa^*p'(\xi)\|^2\\&=\left(\partial_\xi\log\frac{f(\xi)}{C_0}\right)^2f(\xi)+\left(\partial_\xi\log\frac{1-f(\xi)}{C_1}\right)^2(1-f(\xi))=\frac{\partial_\xi f(\xi)^2}{f(\xi)\bigl(1-f(\xi)\bigr)}.\end{aligned}\qquad\square$$

M 上の次の非負関数を考える．

$$h(\xi)=\frac{\mathfrak{g}'(\partial_\xi,\partial_\xi)}{\mathfrak{g}(\partial_\xi,\partial_\xi)}.$$

概十分統計量の定義により，κ が $\delta>0$ に対して δ 概十分となるのは，$h(\xi)$ が M 上で正の最小値を持つことと同値である．[9] により，$\mathfrak{g}(\partial_\xi,\partial_\xi)=\frac{n}{\xi(1-\xi)}$ となる．そして，補題 3.2 により，次の式が得られる．

$$h(\xi)=\frac{\xi(1-\xi)\partial_\xi f(\xi)^2}{nf(\xi)(1-f(\xi))}.\tag{3.1}$$

補題 3.3. $\lim_{\xi\to0}h(\xi)\neq0$ は，$\partial_\xi f(0)\neq0$ のときかつそのときのみに成り立つ．

証明. $p(\omega;\xi)$ は ξ の多項式なので，$f(\xi)$ も多項式である．$m\geq1$ とし，多項式 $g(\xi)$ を，$f(\xi)=\xi^mg(\xi)+f(0)$ かつ $\xi\nmid g(\xi)$ なるものとする．このとき，(3.1) により，次の式が得られる．

$$h(\xi)=\frac{\xi(1-\xi)(m\xi^{m-1}g(\xi)+\xi^m\partial_\xi g(\xi))^2}{n(\xi^mg(\xi)+f(0))(1-f(0)-\xi^mg(\xi))}$$

すべての $\omega\in\Omega$ に対して，密度関数 $p(\omega;\xi)$ は多項式であるので，$\xi\in[0,1]$ に対して定義される．$\omega=(0,0,\ldots,0)$ のとき $p(\omega;0)=1$ であり，$\forall\omega\in\Omega-\{(0,0,\ldots,0)\}$ のとき $p(\omega;0)=0$ であるため，$\kappa((0,0,\ldots,0))=0$ のとき $f(0)=1$，それ以外のとき $f(0)=0$ となる．$f(0)=1$ と仮定する．すると，

$$h(\xi)=-\xi^{m-1}\frac{(1-\xi)(mg(\xi)+\xi\partial_\xi g(\xi))^2}{ng(\xi)(1+\xi^mg(\xi))},$$

$$\frac{(1-\xi)(mg(\xi)+\xi\partial_\xi g(\xi)(\xi))^2}{g(\xi)(1+\xi^mg(\xi))}\to\frac{m^2g(0)}{n}\neq0\quad(\xi\to0),$$

となるので，$m=1$ の場合に限り $\lim_{\xi\to0}h(\xi)\neq0$ となり，$f(0)=1$ の場合の補題の証明が完了する．$f(0)=0$ の場合の証明も同様である．$\square$

同様に，$\lim_{\xi\to1}h(\xi)\neq0$ となるのは，$\partial_\xi f(1)\neq0$ となるときかつそのときのみである．定義により，すべての $0<\xi<1$ に対して，$\partial_\xi f(\xi)=0$ となることから，命題 3.1 が得られる．

命題 3.1 を 2 つの例に適用してみる．

例 3.4. $n=2,3$ の場合の二項分布の概十分統計量 $\kappa\colon\{0,1\}^n\to\{0,1\}$ をすべて書き出す．不要な対称性の議論をしないで済むように，$\kappa((1,1,\ldots,1))=0$ となる κ のみを考える．この場合，κ は $\kappa((0,0,\ldots,0))=1$ を満たす必要がある．実際，そうでない場合は $\kappa((0,0,\ldots,0))=0$ となり，$f(0)=f(1)=1$ となる．すると，Rolle の定理により，$\partial_\xi f$ は $[0,1]$ 上のある点で消えることより，命題 3.1 から，κ は概十分でない．

$n=2$ の場合，κ で $\kappa((0,0))=1$ かつ $\kappa((1,1))=0$ のとき，$f(\xi)=\xi^2+a\xi(1-\xi)$ が成り立つ．ここで $a=\#\kappa^{-1}(0)\cap\{(0,1),(1,0)\}$ である．命題 3.1 で $\partial_\xi f$ を直接計算すると，κ は $a=1$ のときのみ概十分であることが分かる．これは基本的に例 2.2 と一致する．

$n=3$ の場合，κ で $\kappa((0,0,0))=1$ かつ $\kappa((1,1,1))=0$ のとき，$f(\xi)=\xi^3+a_1\xi^2(1-\xi)+a_2\xi(1-\xi)^2$ が成り立つ．ここで，

$$a_1=\#\kappa^{-1}(0)\cap\{(0,1,1),(1,0,1),(1,1,0)\},\ a_2=\#\kappa^{-1}(0)\cap\{(0,0,1),(0,1,0),(1,0,0)\}$$

である．命題 3.1 を用いて $\partial_\xi f$ を直接計算すると，

$$(a_1, a_2) = (0,1), (0,2), (1,1), (1,2), (1,3), (2,1), (2,2), (2,3)$$

の場合にのみ κ が概十分であることがわかる．$(a_1, a_2) = (2,1)$ の場合は本質的に $n = 3$ および $m = 1$ の例 2.2 と同じであることに注意されたい．

二項分布の概十分統計量は，n が大きいほど多くなる．2 つの概十分統計量の合成は概十分なため，$n \geq 3$ の場合，合成 $\{0,1\}^n \to \{0,1\}^3 \to \{0,1\}$ によってバイナリーな概十分統計量を構成できる．ここで，$\{0,1\}^n \to \{0,1\}^3$ は例 2.2 の $m = 3$ のものであり，$\{0,1\}^3 \to \{0,1\}$ は最後の例の 1 つである．

これらの議論により，次の例は概十分ではないことを述べておく．

例 3.5. $n \geq 3$，$0 \leq m \leq n-1$ とする．$\kappa\colon \Omega \to \{0,1\}$ を次のように定義する．

$$\kappa(\omega) = \begin{cases} 0 & a(\omega) \leq m, \\ 1 & a(\omega) > m, \end{cases}$$

ここで，$a(\omega)$ は ω の 1 の数を表す．$m \leq n-2$ の場合，$f(\xi) = \sum_{k=0}^{m} {}_nC_k \xi^k (1-\xi)^{n-k} = (1-\xi)^2 \sum_{k=0}^{m} {}_nC_k \xi^k (1-\xi)^{n-2-k}$ より，$\partial_\xi f(1) = 0$ となる．$m = n-1$ の場合，$f(\xi) = 1 - (1-\xi)^n$ かつ $\partial_\xi f(1) = 0$ となる．したがって，命題 3.1 により，κ は概十分でない．

参考文献

[1] Amari, S., Nagaoka, H.: Methods of Information Geometry. Translated by Daishi Harada. Transl. Math. Monogr. 191. American Mathematical Society, Providence, RI; Oxford University Press, Oxford (2000) https://doi.org/10.1090/mmono/191

[2] Ay, N., Jost, J., Lê, H.V., Schwachhöfer, L.: Parametrized measure models. Bernoulli 24(3), 1692–1725 (2018) https://doi.org/10.3150/16-BEJ910

[3] Ay, N., Jost, J., Lê, H.V., Schwachhöfer, L.: Information geometry and sufficient statistics. Probab. Theory Relat. Fields 162, 327–364 (2015) https://doi.org/10.1007/s00440-014-0574-8

[4] Ay, N., Jost, J., Lê, H.V., Schwachhöfer, L.: Invariant Geometric Structures on Statistical Models. In: Nielsen, F., Barbaresco, F. (eds) Geometric Science of Information.

[5] Ay, N., Jost, J., Lê, H.V., Schwachhöfer, L.: Information Geometry. A Series of Modern Surveys in Mathematics, vol. 64. Springer, Cham (2017) https://doi.org/10.1007/978-3-319-56478-4

[6] Bernardo, J.-M., Smith, A.F.M.: Bayesian Theory. Wiley Ser. Probab. Stat.: Probability and Mathematical Statistics. John Wiley & Sons, Ltd., Chichester (1994) https://doi.org/10.1002/9780470316870

[7] Blackwell, D.: Conditional expectation and unbiased sequential estimation. Ann. Math. Statist. 18(1), 105–110 (1947) https://doi.org/10.1214/aoms/1177730497

[8] Decell, H.P., Jr.: Sufficient, almost sufficient statistics and applications. In: Proceedings of the First World Conference on Mathematics at the Service of Man (Barcelona, 1977), Vol. I, pp. 541–550 Univ. Politec., Barcelona (1980)

[9] Fisher, R.A.: On the mathematical foundations of theoretical statistics. Philos. Trans. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci. containing Papers of a Mathematical or Physical Character 222, 309–368 (1922) http://doi.org/10.1098/rsta.1922.0009

[10] Kolmogorov, A.N.: Unbiased estimates. Izvestiya Akad. Nauk SSSR. Ser. Mat. 14, 303–326 (1950)

[11] Kullback, S.: Information Theory and Statistics. Wiley, New York (1959)

[12] Lê, H.V.: The uniqueness of the Fisher metric as information metric. Ann. Inst. Statist. Math. 69(4), 879–896 (2017) https://doi.org/10.1007/s10463-016-0562-0

[13] Lehmann, E.L., Scheffé, H.: Completeness, similar regions and unbiased estimation. Sankhyā. 10(4), 305–340 (1950)

[14] Neyman, J.: Sur un teorema concernente le considette statistiche sufficient. Giorn. Istit. Ital. Att. 6, 320–334 (1935)

[15] Peters, B.C., Jr., Redner, R., Decell, H.P., Jr: Characterizations of linear sufficient statistics. Sankhyā Ser. A. 40(3), 303–309 (1978)

[16] Rao, C.R.: Information and accuracy attainable in the estimation of statistical parameters. Bull. Calcutta Math. Soc. 37(3), 81–91 (1945) https://doi.org/10.1007/978-1-4612-0919-5_16

[17] Yamaguchi, K., Nozawa, H. On statistics which are almost sufficient from the viewpoint of the Fisher metrics. Info. Geo. (2024). https://doi.org/10.1007/s41884-024-00160-1