KL ダイバージェンスを用いた δ 概十分統計量の特徴づけ

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

十分統計量の定量的な一般化として，著者と野澤による共同研究では，Fisher 計量を用いて定式化された δ 概十分統計量を提案した．本稿では，この δ 概十分統計量を Kullback–Leibler ダイバージェンスを用いて特徴づける．また，反復試行のモデルについて δ 概十分統計量の新しい例を 1 つ示す．

1 導入

統計量によって統計モデルについての情報がどの程度粗くなるかという問題は，統計学において古くから中心的な課題のひとつである．1920 年代には Fisher が，確率分布の族のパラメータ情報を効率的に捉えるため，推定に必要な情報を損なわない統計量の重要性を指摘した [9]．この考えを発展させ，Neyman は十分統計量という概念を定式化した [12]．この十分性の概念は，その後，情報幾何の枠組みの中で幾何学的に解釈されるようになった．甘利–長岡 [1] は，Fisher 計量を用いて，統計量により誘導されるモデルの Fisher 計量が元のモデルと一致するときにその統計量が十分であると定義し，この定義が古典的な Fisher–Neyman 型十分統計量の定義と同値であることを示した．Ay–Jost–Lê–Schwachhöfer [4, 3] はこの枠組みを拡張し，無限次元のモデルや密度関数による記述にも対応できる一般化を行っている．

一方で，Kullback–Leibler [10] は，Kullback–Leibler ダイバージェンスを通じて十分統計量を特徴づけている．特に，モデル分布と観測値との KL ダイバージェンスを最小化することは，最尤推定などの推定法と深く関係しており，統計量による情報の保持や損失を評価する指標としても，KL ダイバージェンスを用いて議論をすることは自然である．

本稿では，これまで Fisher 計量に基づいて定義されていた δ 概十分統計量 [15] の，KL ダイバージェンスによる特徴づけを紹介する [14]，これが δ 概十分統計量による定義と同値であることを確認する．さらに，反復試行に関する例について KL ダイバージェンスを計算した．

2 統計量の十分性とその定量的な一般化

まず，Ay–Jost–Lê–Schwachhöfer [3, 4] による Fisher 計量を用いた十分統計量の定義と，それを定量的に拡張した δ 概十分統計量の定義および具体例について簡単に紹介する．詳細は著者と野澤の論文 [15] を参照されたい．

[3, 4] の意味で，$(M, \Omega, \mathbf{p})$ をパラメータ付き測度モデルとする．すなわち，M は Banach 多様体，Ω は可測空間，$\mathbf{p} : M \to \mathcal{M}(\Omega)$ は，Ω 上の有限測度の空間 $\mathcal{M}(\Omega)$ に値をとる C^1 級写像である．本稿では，情報損失が Fisher 計量によって表されるような 2 次可積分モデルに焦点を当てる．この 2 次可積分性の仮定のもとで，次に定める Fisher 計量が意味を持つ（詳細は [3] を参照）．$(M, \Omega, \mathbf{p})$ が 2 次可積分であり，さらに Ω

上の有限測度 μ_0 によって支配されていると仮定する．つまり，各 $\xi \in M$ に対して，$\mathbf{p}(\xi)$ はある密度関数 $p(\omega;\xi)$ を用いて $\mathbf{p}(\xi) = p(\omega;\xi)\mu_0$ と書けるとする．このとき，M 上の Fisher 計量 $\mathfrak{g}$ は，$v, w \in T_\xi M$ に対して，$\mathfrak{g}(v, w) = \int_\Omega (\partial_v \log p(\cdot;\xi))(\partial_w \log p(\cdot;\xi))\, d\mathbf{p}(\xi)$ で定義される．

$\mathfrak{g}'$ を統計量 κ により誘導されるモデル $(M, \Omega', \kappa_*\mathbf{p})$ の Fisher 計量とする．ここで，統計量とは可測写像 $\kappa : \Omega \to \Omega'$ であり，$\kappa_*\mathbf{p}$ は，写像 $\kappa : \Omega \to \Omega'$ による $\mathbf{p}$ の押し出し（pushforward measure）である．このとき，M 上の接ベクトル v に対して，統計量 κ による（二次の）情報損失は $\mathfrak{g}(v, v) - \mathfrak{g}'(v, v)$ によって定義される [1]．この情報損失は，Fisher 計量に対する単調性定理 [1, 2, 5, 11] により常に非負となり，以下が M の任意の接ベクトル v に対して成り立つ：

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

定義 2.1 (甘利–長岡 [1], Ay–Jost–Lê–Schwachhöfer [3, 4])**.** 情報損失がすべての点，すべての接ベクトルで消えるとき，すなわち，式 (2.1) の不等式が等号として成り立つとき，統計量 κ は十分であると定義される．

また，Ay–Jost–Lê–Schwachhöfer は [2] 中で，この Fisher 計量によって定義された十分統計量が，古典的な Fisher–Neyman 型の十分統計量とある条件下で一致することを示している．

次に，この十分統計量の条件を定量的に緩和した，著者と野澤による δ 概十分統計量の定義を紹介する．

定義 2.2 (δ 概十分統計量 [15])**.** $(M, \Omega, \mathbf{p})$ 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，

$$\delta^2 \mathfrak{g}(v, v) \leq \mathfrak{g}'(v, v) \tag{2.2}$$

が成り立つときにいう．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 (2.1) より，$\mathfrak{g}'(v, v) \leq \mathfrak{g}(v, v)$ が成り立つから，統計量が δ 概十分性を持つとき，

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

が成立する．つまり，$\mathfrak{g}'$ によって誘導される距離は，$\mathfrak{g}$ による距離と双リプシッツ同値である．よって，幾何学的には，Fisher 計量から誘導される Fisher–Rao 距離に関して，δ 概十分統計量はパラメータ空間 M の粗構造を保つ可測写像であるといえる．

著者と野澤 [15] の主結果においては，δ 概十分統計量の特徴づけとして，L^2 ノルムにおける不等式の比較，条件付き確率に関連する写像のリプシッツ性，および古典的な Fisher–Neyman 型の十分統計量との形の比較を記述する条件が，同値なかたちで与えられている．このような特徴づけを通じて，δ 概十分統計量は，古典的な十分統計量が存在しない場面においても，その近似的性質を利用して推定を行う際の手がかりを与えると期待される．

この δ 概十分統計量以外にもいくつかの統計量のクラス，例えばベイズ十分統計量 [6] や線形十分統計量 [13, 8] が研究されていることに留意されたい．

一方で，Kullback–Leibler [10] は，KL ダイバージェンスを通して十分統計量を特徴づけている．ここで，KL ダイバージェンスは，測度 μ に関して可積分かつ正な密度関数 p, q で $\mathbf{p} = p\mu$，$\mathbf{q} = q\mu$ と書ける測度 $\mathbf{p}, \mathbf{q}$ に対して，

$$D_{KL}(\mathbf{p}, \mathbf{q}) = \int_\Omega p \log\left(\frac{p}{q}\right) d\mu$$

により定義される．

定理 2.3 ([10])**.** 任意の統計量 κ と，任意の $\xi, \eta \in M$ に対して，

$$D'_{KL}(\kappa_* \mathbf{p}_\xi, \kappa_* \mathbf{p}_\eta) \leq D_{KL}(\mathbf{p}_\xi, \mathbf{p}_\eta)$$

が成り立ち，等号が成り立つときかつそのときに限り κ は十分統計量である．

また，Fisher 計量と KL ダイバージェンスの関係については次のことがよく知られている ([7] 参照)．$\xi \in M$ を任意にとり，これを固定する．$\eta \in M$ に対して，ξ と η を滑らかなパスで繋ぎ，ξ における M の接ベクトルを v とする．$\Delta\xi = (\xi - \eta) \to 0$ のとき，

$$D_{KL}(\mathbf{p}_\xi, \mathbf{p}_\eta) = \frac{1}{2}\mathfrak{g}(v, v)(\Delta\xi)^2 + o(\|\Delta\xi\|^2). \tag{2.3}$$

3 反復試行モデルにおける δ 概十分統計量の具体例

本節では，δ 概十分統計量が統計モデルにおいてどのように現れるかを示すために，典型的なモデルのひとつとしてコイン投げ（ベルヌーイ試行の反復）に基づくモデルを考える．

例 3.1 (ベルヌーイ試行)**.** $n \in \mathbb{N}$ に対して，

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

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

$$\kappa(\omega_1, \ldots, \omega_n) = (\omega_1, \ldots, \omega_m) \tag{3.1}$$

により定める．このとき，κ は $\sqrt{m/n}$ 概十分である（[15] 参照）．実際，

$$\frac{m}{n}\,\mathfrak{g}(\partial_\xi, \partial_\xi) = \mathfrak{g}'(\partial_\xi, \partial_\xi)$$

が成り立つ．

上の例では，1 次元多様体上のモデルとしてコイン投げの事例を扱った．では次に，0，1，2 のいずれかが出る 3 面ダイスを n 回振った場合のモデルについて考察する．このとき，先の例との違いがどのように現れるかを確認したい．

例 3.2 (3 面ダイス)**.** まず，0，1，2 のいずれかの値をとる 3 面ダイスを n 回振るという状況を，以下のようにパラメータ付き測度モデルによって表現する．

- $M = \{(\xi_0, \xi_1) \in (0, 1)^2 \mid 0 < \xi_0 + \xi_1 < 1\}$,
- $\Omega = \{0, 1, 2\}^n$（ただし数え上げ測度を備える），
- $\omega \in \Omega$ に対して $\mathbf{p}(\xi)(\{\omega\}) = \xi_0^{a(\omega)}\xi_1^{b(\omega)}(1 - \xi_0 - \xi_1)^{n - a(\omega) - b(\omega)}$ で定める．ここで，$a(\omega)$ と $b(\omega)$ は，それぞれ 0 の出る回数と 1 の出る回数を表す．

この例ではコイン投げの場合とは異なり，多様体 M の次元が 2 の場合を考えていることになる．
ここで，$\Omega' = \{0, 1, 2\}$ とし，$\kappa \colon \Omega \to \Omega'$ を次で定める：

$$\kappa(\omega_1, \ldots, \omega_n) = \omega_1 \tag{3.2}$$

すると，この統計量で誘導されるモデル $(M, \Omega', \kappa_* \mathbf{p})$ は次のようになる．

- $M = \{(\xi_0, \xi_1) \in (0,1)^2 \mid 0 < \xi_0 + \xi_1 < 1\}$,
- $\Omega' = \{0,1,2\}$,（ただし数え上げ測度を備える），
- $\omega' \in \Omega'$ に対して，

$$\kappa_* \mathbf{p}(\xi)(\{\omega'\}) = \begin{cases} \xi_0 & (\omega' = 0), \\ \xi_1 & (\omega' = 1), \\ 1 - \xi_0 - \xi_1 & (\omega' = 2). \end{cases}$$

では，ここでオリジナルのモデルの Fisher 計量 $\mathfrak{g}$ と，誘導されたモデルの Fisher 計量 $\mathfrak{g}'$ を比較するために，それぞれの計量を計算する．

この際，計算を扱いやすくするため，確率密度関数をクロネッカーのデルタを用いて書き換える．オリジナルのモデルの確率密度関数は，$\omega = (\omega_1, \omega_2, ..., \omega_n) \in \Omega^n$ に対して，

$$\mathbf{p}(\xi)(\{\omega\}) = \xi_0^{\sum_{i=1}^n \delta_{0,\omega_i}} \xi_1^{\sum_{i=1}^n \delta_{1,\omega_i}} (1 - \xi_0 - \xi_1)^{\sum_{i=1}^n \delta_{2,\omega_i}} \tag{3.3}$$

と書き換えられ，誘導されたモデルの確率密度関数は，$\omega' \in \Omega$ に対して，

$$\kappa_* \mathbf{p}_\xi(\{\omega'\}) = \xi_0^{\delta_{0,\omega'}} \xi_1^{\delta_{1,\omega'}} (1 - \xi_0 - \xi_1)^{\delta_{2,\omega'}} \tag{3.4}$$

と書き換えられる．ここで，記号を簡略化して $\mathbf{p}(\xi)(\{\omega\}) = \mathbf{p}(\omega), \kappa_* \mathbf{p}(\xi)(\{\omega'\}) = \mathbf{p}'(\omega')$ と表すと，$\mathbf{p}(\omega)$ は $\mathbf{p}'(\omega')$ を用いて，

$$\mathbf{p}(\omega) = \Pi_{i=1}^n \mathbf{p}'(\omega_i) \tag{3.5}$$

と書ける．先に誘導されたモデルにおける Fisher 計量を計算すると，

$$\mathfrak{g}'(\partial_{\xi_0}, \partial_{\xi_0}) = \sum_{\omega' \in \{0,1,2\}} \left(\frac{\delta_{0,\omega'}}{\xi_0} - \frac{\delta_{2,\omega'}}{1 - \xi_0 - \xi_1} \right)^2 \cdot \xi_0^{\delta_{0,\omega'}} \xi_1^{\delta_{1,\omega'}} (1 - \xi_0 - \xi_1)^{\delta_{2,\omega'}}$$

となる．

一方，オリジナルのモデルにおける Fisher 計量は，

$$\begin{aligned} \mathfrak{g}(\partial_{\xi_0}, \partial_{\xi_0}) &= \sum_{\omega \in \Omega} \left(\frac{\sum_{j=1}^n \delta_{0,\omega_j}}{\xi_0} - \frac{\sum_{j=1}^n \delta_{2,\omega_j}}{1 - \xi_0 - \xi_1} \right)^2 \mathbf{p}(\omega) \\ &= n \sum_{\omega' \in \{0,1,2\}} \left(\frac{\delta_{0,\omega'}}{\xi_0} - \frac{\delta_{2,\omega'}}{1 - \xi_0 - \xi_1} \right)^2 \cdot \xi_0^{\delta_{0,\omega'}} \xi_1^{\delta_{1,\omega'}} (1 - \xi_0 - \xi_1)^{\delta_{2,\omega'}} \\ &= n \mathfrak{g}'(\partial_{\xi_0}, \partial_{\xi_0}) \end{aligned} \tag{3.6}$$

となり，$\mathfrak{g}(\partial_{\xi_1}, \partial_{\xi_1})$ も同様に計算でき，$\mathfrak{g}(\partial_{\xi_1}, \partial_{\xi_1}) = n\mathfrak{g}'(\partial_{\xi_1}, \partial_{\xi_1})$ がわかる．また，クロスターム $\mathfrak{g}(\partial_{\xi_0}, \partial_{\xi_1})$,

$\mathfrak{g}(\partial_{\xi_1}, \partial_{\xi_0})$ も，$\mathfrak{g}(\partial_{\xi_0}, \partial_{\xi_1}) = n\mathfrak{g}'(\partial_{\xi_0}, \partial_{\xi_1})$, $\mathfrak{g}(\partial_{\xi_1}, \partial_{\xi_0}) = n\mathfrak{g}'(\partial_{\xi_1}, \partial_{\xi_0})$ となる．実際,

$$
\begin{aligned}
\mathfrak{g}(\partial_{\xi_0}, \partial_{\xi_1}) &= \sum_{\omega \in \Omega^n} \left(\partial_{\xi_0} \log \mathbf{p}(\omega)\right) \left(\partial_{\xi_1} \log \mathbf{p}(\omega)\right) \mathbf{p}(\omega) \\
&= \sum_{\omega \in \Omega^n} \left(\partial_{\xi_0} \log \Pi_{i=1}^n \mathbf{p}'(\omega_i)\right) \left(\partial_{\xi_1} \Pi_{i=1}^n \log \mathbf{p}'(\omega_i)\right) \Pi_{i=1}^n \mathbf{p}'(\omega_i) \\
&= \sum_{\omega \in \Omega^n} \left(\partial_{\xi_0} \log \mathbf{p}'(\omega_1) + \cdots + \partial_{\xi_0} \log \mathbf{p}'(\omega_n)\right) \left(\partial_{\xi_1} \log \mathbf{p}'(\omega_1) + \cdots + \partial_{\xi_1} \log \mathbf{p}'(\omega_n)\right) \Pi_{i=1}^n \mathbf{p}'(\omega_i) \\
&= \sum_{\omega \in \Omega^n} \left(\partial_{\xi_0} \log \mathbf{p}'(\omega_1)\right) \left(\partial_{\xi_1} \log \mathbf{p}'(\omega_1)\right) \Pi_{i=1}^n \mathbf{p}'(\omega_i) + \cdots \\
&\quad + \sum_{\omega \in \Omega^n} \left(\partial_{\xi_0} \log \mathbf{p}'(\omega_n)\right) \left(\partial_{\xi_1} \log \mathbf{p}'(\omega_n)\right) \Pi_{i=1}^n \mathbf{p}'(\omega_i) \\
&\qquad + \sum_{\omega \in \Omega^n} \left(\partial_{\xi_0} \log \mathbf{p}'(\omega_1)\right) \left(\sum_{\substack{j=1 \\ j \neq 1}}^n \partial_{\xi_1} \log \mathbf{p}'(\omega_j) \right) \Pi_{i=1}^n \mathbf{p}'(\omega_i) + \cdots \\
&\qquad\quad + \sum_{\omega \in \Omega^n} \left(\partial_{\xi_0} \log \mathbf{p}'(\omega_n)\right) \left(\sum_{\substack{j=1 \\ j \neq n}}^n \partial_{\xi_1} \log \mathbf{p}'(\omega_j) \right) \Pi_{i=1}^n \mathbf{p}'(\omega_i) \\
&= \sum_{\omega \in \Omega^n} \left(\partial_{\xi_0} \log \mathbf{p}'(\omega_1)\right) \left(\partial_{\xi_1} \log \mathbf{p}'(\omega_1)\right) \mathbf{p}'(\omega_1) \Pi_{i=2}^n \mathbf{p}'(\omega_i) + \cdots \\
&\quad + \sum_{\omega \in \Omega^n} \left(\partial_{\xi_0} \log \mathbf{p}'(\omega_n)\right) \left(\partial_{\xi_1} \log \mathbf{p}'(\omega_n)\right) \mathbf{p}'(\omega_n) \Pi_{i=1}^{n-1} \mathbf{p}'(\omega_i) \\
&\qquad + \sum_{\omega \in \Omega^n} \left(\partial_{\xi_0} \mathbf{p}'(\omega_1)\right) \left(\sum_{\substack{j=1 \\ j \neq 1}}^n \partial_{\xi_1} \log \mathbf{p}'(\omega_j) \right) \Pi_{i=2}^n \mathbf{p}'(\omega_i) + \cdots \\
&\qquad\quad + \sum_{\omega \in \Omega^n} \left(\partial_{\xi_0} \mathbf{p}'(\omega_n)\right) \left(\partial_{\xi_1} \sum_{\substack{j=1 \\ j \neq n}}^n \log \mathbf{p}'(\omega_j) \right) \Pi_{i=1}^{n-1} \mathbf{p}'(\omega_i) \\
&= n \sum_{\omega' \in \Omega} \left(\partial_{\xi_0} \log \mathbf{p}'(\omega')\right) \left(\partial_{\xi_1} \log \mathbf{p}'(\omega')\right) \mathbf{p}'(\omega') \\
&\quad + \partial_{\xi_0} \sum_{\omega_1 \in \Omega} \mathbf{p}'(\omega_1) \sum_{(\omega_2, \ldots, \omega_n)} \left(\left(\sum_{\substack{j=1 \\ j \neq 1}}^n \partial_{\xi_1} \log \mathbf{p}'(\omega_j) \right) \Pi_{i=2}^n \mathbf{p}'(\omega_i) \right) + \cdots \\
&\qquad + \partial_{\xi_0} \sum_{\omega_n \in \Omega} \mathbf{p}'(\omega_n) \sum_{(\omega_1, \ldots, \omega_{n-1})} \left(\left(\partial_{\xi_1} \sum_{\substack{j=1 \\ j \neq n}}^n \log \mathbf{p}'(\omega_j) \right) \right) \Pi_{i=1}^{n-1} \mathbf{p}'(\omega_i) \\
&= n\mathfrak{g}'(\partial_0, \partial_1).
\end{aligned}
$$

よって，上のクロスタームの計算と (3.6) より，統計量 (3.2) は，$\sqrt{1/n}$ 概十分である．またこれらの計算結果より，コイン投げの場合と変わらず，$1 \leq m \leq n$ に対して，$\Omega' = \{0, 1\}^m$ とし，$\kappa \colon \Omega \to \Omega'$ を (3.1) で定めると，κ は $\sqrt{m/n}$ 概十分であることが分かる．

4 KL ダイバージェンスを用いた δ 概十分統計量と具体例

本節では，これまで Fisher 計量に基づいて考えていた δ 概十分統計量を，KL ダイバージェンスの視点から考察する．また，第 2 節で扱った具体例の KL ダイバージェンスを計算する．

命題 4.1 ([14])**.** $(M, \Omega, \mathbf{p})$ をパラメータ付き測度モデルとし，その Fisher 計量が M の各点でリーマン計量であると仮定する．十分に近い $\xi, \eta \in M$ に対して，$\delta^2 D_{KL}(\mathbf{p}_\xi, \mathbf{p}_\eta) \leq D'_{KL}(\kappa_* \mathbf{p}_\xi, \kappa_* \mathbf{p}_\eta)$ であることと，任意の $v \in T_\xi M$ に対して $\delta^2 \mathfrak{g}(v, v) \leq \mathfrak{g}'(v, v)$ であることは同値である．

証明. $\xi \neq \eta$ なる $\eta \in M$ に対して，ξ と η を滑らかなパスで繋ぎ，ξ における M の接ベクトルを v とする．$\Delta\xi = (\xi - \eta) \to 0$ のとき (2.3) より，

$$D_{KL}(\mathbf{p}_\xi, \mathbf{p}_\eta) = \frac{1}{2}\mathfrak{g}(v, v)(\Delta\xi)^2 + o(\|\Delta\xi\|^2).$$

これを $\epsilon - \delta$ 論法によって書き換えると，

$$\forall \epsilon > 0, \exists \delta > 0 \text{ s.t. } 0 < |\xi - \eta| < \delta \Rightarrow \left| \frac{D_{KL}(\mathbf{p}_\xi, \mathbf{p}_\eta) - \frac{1}{2}\mathfrak{g}(v, v)(\Delta\xi)^2}{\|\Delta\xi\|^2} \right| < \epsilon.$$

であり，さらに，

$$-\epsilon\|\Delta\xi\|^2 + \frac{1}{2}\mathfrak{g}(v, v)(\Delta\xi)^2 < D_{KL}(\mathbf{p}_\xi, \mathbf{p}_\eta) < \epsilon\|\Delta\xi\|^2 + \frac{1}{2}\mathfrak{g}(v, v)(\Delta\xi)^2. \tag{4.1}$$

と書ける．一方で，統計量で誘導されたモデル上でも同様の議論により，

$$-\epsilon\|\Delta\xi\|^2 + \frac{1}{2}\mathfrak{g}'(v, v)(\Delta\xi)^2 < D'_{KL}(\kappa_* \mathbf{p}_\xi, \kappa_* \mathbf{p}_\eta) < \epsilon\|\Delta\xi\|^2 + \frac{1}{2}\mathfrak{g}'(v, v)(\Delta\xi)^2. \tag{4.2}$$

よって，$(4.2) - \delta^2(4.1)$ より，

$$\begin{aligned} &-\epsilon\|\Delta\xi\|^2 - \delta^2\epsilon\|\Delta\xi\|^2 + \frac{1}{2}\mathfrak{g}'(v, v)(\Delta\xi)^2 - \frac{1}{2}\delta^2\mathfrak{g}(v, v)(\Delta\xi)^2 < D'_{KL}(\kappa_* \mathbf{p}_\xi, \kappa_* \mathbf{p}_\eta) - \delta^2 D_{KL}(\mathbf{p}_\xi, \mathbf{p}_\eta) \\ &\qquad < \epsilon\|\Delta\xi\|^2 + \delta^2\epsilon\|\Delta\xi\|^2 + \frac{1}{2}\mathfrak{g}'(v, v)(\Delta\xi)^2 - \frac{1}{2}\delta^2\mathfrak{g}(v, v)(\Delta\xi)^2. \end{aligned} \tag{4.3}$$

ここで，$\delta^2 \mathfrak{g}(v, v) \leq \mathfrak{g}'(v, v)$ が成り立つと仮定すると，仮定より $\mathfrak{g}(v, v) > 0$ のため ϵ を十分小さく取ることによって，$D'_{KL}(\kappa_* \mathbf{p}_\xi, \kappa_* \mathbf{p}_\eta) - \delta^2 D_{KL}(\mathbf{p}_\xi, \mathbf{p}_\eta) > 0$ が言える．逆も同様に示される． □

次に，コイン投げのモデルが，KL ダイバージェンス版でも δ 概十分統計量の例になっていることを確認する．

例 4.2. $0 < m < n$ として, まず n 回コインを投げる状況で，KL ダイバージェンスを計算する．

$$\begin{aligned} D_{KL}(\mathbf{p}_\xi, \mathbf{p}_\eta) &= \sum_{k=0}^{n} \binom{n}{k} \xi^k (1-\xi)^{n-k} \log \frac{\xi^k (1-\xi)^{n-k}}{\eta^k (1-\eta)^{n-k}} \\ &= \sum_{k=0}^{n} \binom{n}{k} \xi^k (1-\xi)^{n-k}\, k \log \frac{\xi}{\eta} + \sum_{k=0}^{n} \binom{n}{k} \xi^k (1-\xi)^{n-k} (n-k) \log \frac{1-\xi}{1-\eta} \\ &= \log \frac{\xi}{\eta} \left(\sum_{k=0}^{n} k \binom{n}{k} \xi^k (1-\xi)^{n-k} \right) + \log \frac{1-\xi}{1-\eta} \left(\sum_{k=0}^{n} (n-k) \binom{n}{k} \xi^k (1-\xi)^{n-k} \right) \\ &= n\xi \log \frac{\xi}{\eta} + n(1-\xi) \log \frac{1-\xi}{1-\eta} \\ &= n \left(\xi \log \frac{\xi}{\eta} + (1-\xi) \log \frac{1-\xi}{1-\eta} \right). \end{aligned}$$

よって，m 回コインを投げるモデルについても同様に KL ダイバージェンスが計算でき，$D'_{KL}(\mathbf{p}_\xi, \mathbf{p}_\eta) = \sqrt{m/n} D_{KL}(\mathbf{p}_\xi, \mathbf{p}_\eta)$ であるので，(3.1) の統計量を考える場合，その統計量が $\sqrt{\frac{m}{n}}$ 概十分であることがこの計算と命題 4.1 とよりわかる．

謝辞

本研究は，JSPS 科研費 25KJ2213 の支援を受けたものである．

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.: Information geometry and sufficient statistics. Probab. Theory Relat. Fields 162, 327–364 (2015) https://doi.org/10.1007/s00440-014-0574-8

[3] 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

[4] 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

[5] 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.

[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] Calin, O., Udrişte, C.: Geometric Modeling in Probability and Statistics. Springer, Cham (2014) https://doi.org/10.1007/978-3-319-07779-6

[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] Kullback, S., Leibler, R.A.: On Information and Sufficiency. Ann. Math. Statist. **22**(1), 79–86 (1951) https://doi.org/10.1214/aoms/1177729694

[11] 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

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

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

[14] Yamaguchi,K. Quantitatively almost sufficient statistics and the Kullback–Leibler divergence, In preparation

[15] 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