日本統計学会誌
第37巻，第2号，2008年3月
329頁～342頁

非正則推定における情報量の概念とその役割

赤 平 昌 文*

The concept of amount of information and its role in non-regular estimation

Masafumi Akahira*

統計的推定論において，推定量のリスクに関する情報不等式を用いる際に，情報量は重要な役割を果している．また，高次漸近理論において，情報量損失（または漸近欠損量）を求めることによって統計量（または推定量）の高次の漸近的比較を試みることができる（Akahira, 1986；赤平，2006）．本稿では，特に正則条件が必ずしも成り立たないような場合の非正則推定に焦点を当てて，情報量の概念とその役割について考える．実際に，いくつかの情報量等を導入して相互関係について述べるとともに，それに基づいて推定量の一致性，（漸近）効率等について論じる．

In statistical estimation, amounts of information play an important part in using information inequalities for the risk of estimators. In the higher order asymptotic theory, it is possible to try the asymptotic comparison among statistics (or estimators) up to the higher order, through the loss of information (or asymptotic deficiency) associated with them (Akahira, 1986, 2006). In this paper, we focus our subject on the non-regular estimation under which the regularity conditions do not necessarily hold, and consider the concept of amount of information and its role. Indeed, we introduce some amounts of information, investigate their mutual relationships and discuss the consistency, (asymptotic) efficiency, etc. of estimators based on them.

Key Words and Phrases: Fisher information, generalized amount of information, loss of information, total variation, $\{c_n\}$・consistency, asymptotic efficiency

1. はじめに

統計的推定論において，情報量の概念が果す役割は大きいが，通常の Fisher 情報量，Kullback-Leibler 情報量等が有用なのは正則な場合に限られる．特に，適当な正則条件が成り立てば，Cramér-Rao の不等式から不偏推定量の分散の下界が Fisher 情報量の逆数として与えられる．しかし，非正則な場合の1つとして Fisher 情報量が無限大になる場合が考えられ，その場合には不偏推定量の局所分散の下限が0になることが示される．また，確率変数 X が一様分布 $U(\theta-(1/2), \theta+(1/2))$ に従うときに，X がもつ θ に関する Fisher 情報量は0となってしまい，無意味になる．一方，その一様分布からの無作為標本に基づく θ の Pitman 推定量（最良位置共変推定量）は範囲の中央（midrange）になるが（Ibragimov and Has'minskii, 1981），それのもつ θ に関する何らかの情報量を考えて，元の標本に基づく情報量と比較することができるだろうか．さらに，2つの統計量が1次の次数（order）では同じ情報量を持っていても，次の次数では異なることも多いので，その次数での情報量の漸近的な差，すなわち情報量損失が意味をもつ．実際，密度が滑らかでないような非正則の場合については，Fisher（1925,

* 第12回日本統計学賞受賞者．筑波大学大学院数理物質科学研究科：〒305-8571 つくば市天王台 1-1-1

1934）が両側指数分布の場合に最尤推定量の情報量損失を計算し，その漸近的な次数が $\sqrt{n}$ であり，これは正則な場合に定数の次数になることと比較して特異であると指摘している（Akahira and Takeuchi, 1995, Chap. 4；福水・栗木 他 2004，補論 B）．また，密度の台（support）が有界になるような非正則な場合については，Neyman が，「変異係数が従う確率法則の性質について」という論文（*Comp. Rend.* **183**, 1926, pp. 107–109）をフランス科学アカデミーに提出し，その中で次のような非正則問題を考察した．いま，$X_1,\cdots,X_n$ が $f_0(x)>0$ $(a<x<b)$; $f_0(x)=0$ （$x\leq a$ または $x\geq b$）となる密度 f_0 をもつ分布からの無作為標本とするとき，$\bar{X}$ を標本平均，S^2 を標本分散とする．このとき，$\nu:=S/\bar{X}$ を変異係数といい，これは人体計測学でよく用いられる統計量で，それの原点周りの k の積率の値が有限確定になるための必要十分条件を求めた（安藤，1985）．さらに，D. Bernoulli, L. Euler のような高名な先人達も上記の密度 $f_0(x-\theta)$ の特別な場合に，位置母数 θ の推定問題を考えていた（第5節参照；大谷内・赤平，2007）．

本稿では，上記のことを含めて，非正則推定において情報量の概念とそれが果たす役割について考える．また，全変動による情報量を含むいくつかの情報量を導入して，推定量の一致性とその次数，(漸近)効率等について論じる．

2. Fisher 情報量が無限大に発散する場合

確率ベクトル (X, Y) の（σ-有限測度に関する）同時確率密度関数（j.p.d.f.）を $f_{X,Y}^{\theta}(x, y)$ とする．ただし，$\theta\in\Theta$ で Θ は $\boldsymbol{R}^1$ の開区間とする．いま，$Y=y$ を与えたときの X の条件付 p.d.f. を $f_{X|Y}^{\theta}(x|y)$ とし，Y の周辺 p.d.f. を $f_Y^{\theta}(y)$ とする．また，$f_{X,Y}^{\theta}, f_{X|Y}^{\theta}, f_Y^{\theta}$ は θ に関して微分可能として，それぞれの Fisher 情報量を $I_{X,Y}(\theta), I_{X|Y}(\theta), I_Y(\theta)$ とすれば，

$$\log f_{X,Y}^{\theta}(x, y)=\log f_{X|Y}^{\theta}(x|y)+\log f_Y^{\theta}(y) \quad a.e.$$

より

$$I_{X,Y}(\theta)=E_{\theta}^{Y}[I_{X|Y}(\theta)]+I_Y(\theta)$$

になる（Kullback, 1959）．ただし，$I_{X,Y}(\theta):=E_{\theta}[\{(\partial/\partial\theta)\log f_{X,Y}^{\theta}(X, Y)\}^2]$ とし，$I_{X|Y}(\theta)$, $I_Y(\theta)$ も同様に定義し，$E_{\theta}^{Y}(\cdot)$ は f_Y^{θ} に関する期待値とする．ここで，$J_{\theta}(Y):=I_{X|Y}(\theta)$ とおいて，次の条件（A1）～(A3）を仮定する．

（A1） 任意の $\theta\in\Theta$ と任意の $y\in\mathcal{Y}$ について，$0<J_{\theta}(y)<\infty$ であり，ある $\theta_0\in\Theta$ について

$$E_{\theta_0}^{Y}[J_{\theta_0}(Y)]=\infty$$

である．

（A2） Y の標本空間 $\mathcal{Y}$ の可測集合列 $\{B_n\}_{n=1}^{\infty}$ が存在して，各 $n=1, 2,\cdots$ と各 $\theta\in\Theta$ について $b_n:=E_{\theta}^{Y}[J_{\theta_0}(Y)\chi_{B_n}(Y)]<\infty$ で，b_n は θ に無関係であり，$b_n\to\infty\,(n\to\infty)$ である．ただし，$\chi_{B_n}(\cdot)$ は B_n の定義関数とする．

（A3） 任意の $y\in\mathcal{Y}$ について，θ の推定量 $\hat{\theta}_y(X)$ が存在して，任意の $\theta\in\Theta$ について

$$E_{\theta}[\hat{\theta}_Y(X)\,|\,Y=y]=\theta,$$
$$V_{\theta}[\hat{\theta}_Y(X)\,|\,Y=y]=1/J_{\theta}(y)$$

である．ただし，$E_{\theta}[\cdot|Y=y]$, $V_{\theta}[\cdot|Y=y]$ はそれぞれ y を与えたときの条件付期待値，条件付分散とする．

ここで，$\mathcal{U}$ を (X, Y) に基づく θ の不偏推定量全体の集合とし，また，条件（A1）より $I_{X,Y}(\theta_0)=\infty$ となることに注意すれば，次のようなことが成り立つ（Akahira and Takeuchi, 1985, 1995）．

定理 2.1 条件（A1）～(A3）の下で

$$\inf_{\hat{\theta}\in\mathcal{U}} V_{\theta_0}(\hat{\theta}(X, Y))=0$$

である．

例 2.1 線形モデル $X=\theta Y+U(\theta\in\boldsymbol{R}^1)$ において，Y と U は互いに独立な確率変数で，Y の分布は θ に無関係とし，U の p.d.f. は $p(\cdot)$ で既知とし，$\lim_{|u|\to\infty}p(u)=0$ とする．また

$$I_U:=\int\frac{\{p'(u)\}^2}{p(u)}du<\infty \qquad (p \text{ に関する Fisher 情報量})$$

とする．いま，上のモデルにおいて，$(Y_i, U_i)(i=1,\cdots, n)$ を互いに独立に同分布に従う確率ベクトルとすれば，$\boldsymbol{Y}:=(Y_1,\cdots, Y_n)$ とおくと

$$I_Y(\theta)=0, \quad E^Y[J_\theta(\boldsymbol{Y})]=nE(Y_1^2)I_U$$

となり，$E(Y_1^2)=\infty$ とすれば，$\theta=0$ のとき

$$\inf_{\hat{\theta}\in\mathcal{U}} V_0(\hat{\theta})=0$$

となることが示される．

例 2.2 Y を自由度 1 のカイ 2 乗分布に従う未観測の確率変数とし，また，Y を与えたときに，$X_1,\cdots, X_n$ は互いに独立にいずれも正規分布 $N(\theta, Y)(\theta\in\boldsymbol{R}^1)$ に従うとする．ただし，$n\geq 5$ とする．ここで，$\boldsymbol{X}:=(X_1,\cdots, X_n)$ とおくと $I_{X,Y}(\theta)=\infty$ になり，$\boldsymbol{X}$ に基づく θ の不偏推定量全体の集合を $\mathcal{U}_0$ とすれば，$\theta=0$ のとき

$$\inf_{\hat{\theta}\in\mathcal{U}_0} V_0(\hat{\theta})=0$$

になることが示される．

3. 一般情報量

一般に，標本空間 $(\mathcal{X}, \mathcal{B})$ 上の確率分布族を $\mathcal{P}:=\{P_\theta: \theta\in\Theta\}$ とし，Θ を母数空間とする．ここで，ある σ-有限測度 μ によって $\mathcal{P}$ が支配されているとき，すなわち，任意の $\theta\in\Theta$ について P_θ は μ に関して絶対連続であるとき

$$f(x, \theta):=\frac{dP_\theta}{d\mu}(x)$$

と表わし，これを（μ に関する）p.d.f. という．いま，確率変数 X が p.d.f. f をもつ分布に従うならば，任意の $\theta_1, \theta_2\in\Theta$ について

$$K_X(\theta_1, \theta_2):=E_{\theta_1}\left[\log\frac{f(X, \theta_1)}{f(X, \theta_2)}\right]=\int_{\mathcal{X}}\left\{\log\frac{f(x, \theta_1)}{f(x, \theta_2)}\right\}f(x, \theta_1)d\mu(x)$$

によって，K_X を定義するとき，これを Kullback-Leibler（K-L）情報量という（Kullback, 1959;

Pardo, 2006)．特に，Θを$\boldsymbol{R}^1$の開区間とすれば，適当な正則条件の下で，$\Delta\to 0$のとき

$$K_X(\theta,\theta+\Delta)=\frac{\Delta^2}{2}I_X(\theta)+o(\Delta^2)$$

になる．ただし，$I_X(\theta)$はFisher情報量，すなわち$I_X(\theta)=E_\theta[\{(\partial/\partial\theta)\log f(X,\theta)\}^2]$とする．

次に，任意の$\theta_1,\theta_2\in\Theta$について

$$H_X(\theta_1,\theta_2):=\frac{1}{\sqrt{2}}\left[\int_{\mathcal{X}}\{\sqrt{f(x,\theta_1)}-\sqrt{f(x,\theta_2)}\}^2d\mu(x)\right]^{1/2}$$

によってH_Xを定義するとき，これをHellinger距離という．また，任意の$\theta_1,\theta_2\in\Theta$について

$$a_X(\theta_1,\theta_2):=\int_{\mathcal{X}}\sqrt{f(x,\theta_1)f(x,\theta_2)}\,d\mu(x)$$

によってa_Xを定義するとき，これを類似度（affinity）という（Matusita, 1955)．さらに，任意の$\theta_1,\theta_2\in\Theta$について

$$d_X(\theta_1,\theta_2):=\sup_{B\in\mathcal{B}}|P_{\theta_1}(B)-P_{\theta_2}(B)|$$

によってd_Xを定義するとき，これを全変動（total variation）による距離という．このとき

$$d_X(\theta_1,\theta_2)=\frac{1}{2}\int_{\mathcal{X}}|f(x,\theta_1)-f(x,\theta_2)|d\mu(x)$$

と表わすこともできる．上記のH, a, dの間には次の関係式

$$H_X^2(\theta_1,\theta_2)=1-a_X(\theta_1,\theta_2),$$
$$H_X^2(\theta_1,\theta_2)\le d_X(\theta_1,\theta_2)\le\{H_X^2(\theta_1,\theta_2)(2-H_X^2(\theta_1,\theta_2))\}^{1/2}$$

が成り立つ（Strasser, 1985; Akahira and Takeuchi, 1995)．

上記に挙げた情報量，距離は，いずれも分布$P_{\theta_1}, P_{\theta_2}$の間の大きさの尺度を表わし，非負値をとり2つの分布が近ければ小さく，離れていれば大きくなる．なお，類似度については，$0\le a_X(\theta_1,\theta_2)\le 1$で，その値の意味は距離とは逆になることに注意．また，これらのうち，Fisher情報量，K-L情報量は，fの台$\{x|f(x,\theta)>0\}$がθに依存するような非正則分布族に対してはあまり使えない．また，他の距離は非正則な場合にも通用するが，正則な場合のFisher情報量との関係が必ずしも明瞭でない．そこで，上記のことを踏まえて，Akahira (1996)は，Rényi型の情報量として任意の$\theta_1,\theta_2\in\Theta$について

$$I_X^{(\alpha)}(\theta_1,\theta_2):=-\frac{8}{1-\alpha^2}\log\int_{\mathcal{X}}f(x,\theta_1)^{(1-\alpha)/2}f(x,\theta_2)^{(1+\alpha)/2}d\mu(x) \tag{3.1}$$

によって，一般情報量（generalized amount of information）を定義した（Rényi, 1961)．ただし，$|\alpha|<1$とする．特に，Θを$\boldsymbol{R}^1$の開区間とすれば，適当な正則条件の下で，任意の$\alpha(|\alpha|<1)$，任意の$\theta\in\Theta$について，$\Delta\to 0$のとき

$$I_X^{(\alpha)}(\theta,\theta+\Delta)=I_X(\theta)\Delta^2+o(\Delta^2)$$

になる．ただし，I_XはFisher情報量とする．

いま，$X_1, X_2,\cdots, X_n,\cdots$を互いに独立にいずれもp.d.f. $f(x,\theta)(\theta\in\Theta)$をもつ分布に従う確率

変数列とする．このとき，$\boldsymbol{X}=(X_1, \cdots, X_n)$ の一般情報量を $I_{\boldsymbol{X}}^{(\alpha)}$ とすれば，任意の $\theta_1, \theta_2 \in \Theta$ について

$$I_{\boldsymbol{X}}^{(\alpha)}(\theta_1, \theta_2)=nI_{X_1}^{(\alpha)}(\theta_1, \theta_2) \quad (\text{加法性})$$

が成り立つ．一般に，統計量 $S_n:=S_n(\boldsymbol{X})$ の一般情報量も(3.1)と同様に定義して，それを $I_{S_n}^{(\alpha)}(\cdot,\cdot)$ と表わせば，適当な正則条件の下で

$$I_{S_n}^{(\alpha)}(\theta_1, \theta_2) \leq I_{\boldsymbol{X}}^{(\alpha)}(\theta_1, \theta_2) \tag{3.2}$$

が成り立ち，等号は S_n が (θ_1, θ_2) に対して十分統計量のときに限り成り立つ．

いま，特に $\alpha=0$ の場合，すなわち一般情報量

$$I_{\boldsymbol{X}}^{(0)}(\theta_1, \theta_2)=-8\log\int_{\chi}\sqrt{f(x, \theta_1)f(x, \theta_2)}\,d\mu(x) \tag{3.3}$$

を用いて具体的な分布族について考察する（Akahira and Takuchi, 1991a; LeCam, 1990）．ここで，(3.3) の右辺の積分が類似度 $a_X(\theta_1, \theta_2)$ であることに注意．

例 3.1 $X_1, \cdots, X_n$ が互いに独立にいずれも正規分布 $N(\mu_j, \sigma_j^2)$ に従う確率変数とする $(j=1, 2)$．このとき，$\theta_j=(\mu_j, \sigma_j^2)\,(j=1, 2)$ とすれば，各 $i=1, \cdots, n$ について

$$I_{X_i}^{(0)}(\theta_1, \theta_2)=4\log\left(\frac{\sigma_1^2+\sigma_2^2}{2\sigma_1\sigma_2}\right)+\frac{2(\mu_1-\mu_2)^2}{\sigma_1^2+\sigma_2^2} \tag{3.4}$$

となるから，$\boldsymbol{X}:=(X_1, \cdots, X_n)$ について

$$I_{\boldsymbol{X}}^{(0)}(\theta_1, \theta_2)=nI_{X_1}^{(0)}(\theta_1, \theta_2) \tag{3.5}$$

になる．一方，$T_1:=\bar{X}=(1/n)\sum_{i=1}^n X_i$, $T_2:=\sum_{i=1}^n(X_i-\bar{X})^2$ とすれば，

$$I_{T_1}^{(0)}(\theta_1, \theta_2)=4\log\left(\frac{\sigma_1^2+\sigma_2^2}{2\sigma_1\sigma_2}\right)+\frac{2n(\mu_1-\mu_2)^2}{\sigma_1^2+\sigma_2^2}, \tag{3.6}$$

$$I_{T_2}^{(0)}(\theta_1, \theta_2)=4(n-1)\log\left(\frac{\sigma_1^2+\sigma_2^2}{2\sigma_1\sigma_2}\right) \tag{3.7}$$

となる．ここで，T_1 と T_2 の独立性および (3.4)～(3.7) より

$$I_{T_1,T_2}^{(0)}(\theta_1, \theta_2)=I_{T_1}^{(0)}(\theta_1, \theta_2)+I_{T_2}^{(0)}(\theta_1, \theta_2)=I_{\boldsymbol{X}}^{(0)}(\theta_1, \theta_2) \tag{3.8}$$

になる．よって，(3.2)，(3.8) より (T_1, T_2) が (θ_1, θ_2) に対して十分統計量になることが分かる．

例 3.2 $X_1, \cdots, X_n$ が互いに独立に，いずれも p.d.f.

$$f(x, \xi_j, \tau_j)=\begin{cases}\dfrac{1}{\tau_j}\exp\left(-\dfrac{x-\xi_j}{\tau_j}\right) & (x>\xi_j),\\ 0 & (x\leq\xi_j)\end{cases}$$

をもつ分布に従う確率変数とする $(j=1, 2)$．ただし，$-\infty<\xi_j<\infty, \tau_j>0$ $(j=1, 2)$ とする．この場合に，各 j について $f(\cdot, \xi_j, \tau_j)$ の台は ξ_j に依存することに注意．いま，$\xi_1<\xi_2$ とする．

このとき，$\theta_j=(\xi_j, \tau_j)(j=1, 2)$ とすれば，各 $i=1,\cdots, n$ について

$$I_{X_i}^{(0)}(\theta_1, \theta_2)=4\log\frac{(\tau_1+\tau_2)^2}{4\tau_1\tau_2}+\frac{4(\xi_2-\xi_1)}{\tau_1} \tag{3.9}$$

となるから，$\boldsymbol{X}:=(X_1,\cdots, X_n)$ について

$$I_{\boldsymbol{X}}^{(0)}(\theta_1, \theta_2)=nI_{X_1}^{(0)}(\theta_1, \theta_2) \tag{3.10}$$

になる．一方，$T_1:=\min_{1\le i\le n}X_i$, $T_2:=\bar{X}-\min_{1\le i\le n}X_i$ とすれば

$$I_{T_1}^{(0)}(\theta_1, \theta_2)=4\log\frac{(\tau_1+\tau_2)^2}{4\tau_1\tau_2}+4n\frac{\xi_2-\xi_1}{\tau_1}, \tag{3.11}$$

$$I_{T_2}^{(0)}(\theta_1, \theta_2)=4(n-1)\log\frac{(\tau_1+\tau_2)^2}{4\tau_1\tau_2} \tag{3.12}$$

となる．ここで，T_1 と T_2 の独立性および (3.9)〜(3.12) より

$$I_{T_1, T_2}^{(0)}(\theta_1, \theta_2)=I_{T_1}^{(0)}(\theta_1, \theta_2)+I_{T_2}^{(0)}(\theta_1, \theta_2)=I_{\boldsymbol{X}}^{(0)}(\theta_1, \theta_2) \tag{3.13}$$

を得る．よって，(3.2)，(3.13) より (T_1, T_2) が (θ_1, θ_2) に対して十分統計量になることが分かる．

4. 非正則分布族における統計量の2次の一般情報量損失

まず，$X_1, X_2,\cdots, X_n,\cdots$ を互いに独立にいずれも（Lebesgue 測度に関する）p.d.f. $f(x, \theta)$ $(\theta\in\Theta)$ をもつ分布に従う実確率変数列とする．ただし，$\Theta=\boldsymbol{R}^1$ とする．このとき θ が位置母数，すなわち $f(x, \theta)=f_0(x-\theta)$ の場合を考える．また f_0 に次の条件を仮定する．

(B1) $f_0(x)>0$ $(a<x<b)$; $f_0(x)=0$ $(x\le a, x\ge b)$．ただし a，b は有限とする．
(B2) $f_0(x)$ は開区間 (a, b) において2回連続微分可能である．
(B3) $\lim_{x\to a+0}f_0(x)=\lim_{x\to b-0}f_0(x)=c$, $\lim_{x\to b-0}f_0'(x)=-\lim_{x\to a+0}f_0'(x)=h$ である．ただし，c は正の定数で，h は定数とする．
(B4) $0<I_0:=\int_a^b\{f_0'(x)\}^2/f_0(x)dx<\infty$.

ここで

$$I:=-\int_a^b\frac{d^2\log f_0(x)}{dx^2}f_0(x)dx$$

とおくと，条件 (B1)〜(B4) より $I-I_0=-2h$ となり，また $0<I<\infty$ になる．なお，上記の条件の下では，推定量の一致性の次数は n であることが知られている．このとき $\boldsymbol{X}:=(X_1,\cdots, X_n)$ の一般情報量を求めると，次のようになる (Akahira, 1996).

定理 4.1 条件 (B1)〜(B4) の下で，$|\alpha|<1$ について，$\Delta\to 0$ とすると

$$I_{\boldsymbol{X}}^{(\alpha)}(\theta, \theta+\Delta)=\frac{1}{1-\alpha^2}[8cn|\Delta|+\{4c^2-2h+I-\alpha^2(2h+I)\}n\Delta^2]+o(n\Delta^2) \tag{4.1}$$

が成り立つ．

次に，極値統計量 $\bar{\theta}$ と $\underline{\theta}$ を $\bar{\theta}:=\min_{1\le i\le n}X_i-a$, $\underline{\theta}:=\max_{1\le i\le n}X_i-b$ とし，$Z_1(\theta):=-(1/\sqrt{n})\sum_{i=1}^{n}f_0'(X_i-\theta)/f_0(X_i-\theta)$ $(\underline{\theta}<\theta<\bar{\theta})$ とする．また $\hat{\theta}^*=(\underline{\theta}+\bar{\theta})/2$ とおくと $\hat{\theta}^*$ は θ の一致推定量になる．さらに $Z_1^*:=Z_1(\hat{\theta}^*)$ とおくと，Z_1^* は漸近補助統計量になる．

一般に，統計量 $T_n:=T_n(\boldsymbol{X})$ の 2 次の一般情報量損失を，任意の $\alpha(|\alpha|<1)$ に対して

$$L_n^{(\alpha)}(T_n):=\frac{1}{n\Delta^2}\{I_X^{(\alpha)}(\theta,\theta+\Delta)-I_{T_n}^{(\alpha)}(\theta,\theta+\Delta)\}+o(1)\quad(\Delta\to 0)$$

で定義する．ただし，$\Delta=O(1/n)$ とする．このとき，統計量 $T_n^*=(Z_1^*/(\sqrt{n}I_0),\underline{\theta},\bar{\theta})$ の一般情報量を求めると，(4.2)の右辺と同じになり，次の結果を得る(Akahira, 1996).

定理 4.2　条件 (B1)〜(B4) の下で，$\Delta=O(1/n)$ とすれば，統計量 T_n^* の 2 次の一般情報量損失は，任意の $\alpha(|\alpha|<1)$ に対して，$L_n^{(\alpha)}(T_n^*)=o(1)(n\to\infty)$ であり，また，統計量 $(\underline{\theta},\bar{\theta})$ の 2 次の一般情報量損失は，任意の $\alpha(|\alpha|<1)$ に対して，$L_n^{(\alpha)}(\underline{\theta},\bar{\theta})=I_0+o(1)(n\to\infty)$ である．

定理 4.2 から，任意の $\alpha(|\alpha|<1)$ に対して，$I_X^{(\alpha)}(\theta,\theta+\Delta)-I_{T_n^*}^{(\alpha)}(\theta,\theta+\Delta)=o(1/n)$ $(n\to\infty)$ になり，統計量 T_n^* の一般情報量損失は 2 次の次数まで，すなわち $o(1/n)$ まで 0 になる．このことは，一方向型分布族に対して T_n^* が 2 次の漸近十分統計量になるという結果とも符合している (Akahira, 1993)．また，定理 4.2 の結果が α について無関係であることに注意．さらに，定理 4.2 から p.d.f. $f_0(x-\theta)$ の台の両端点での情報量を与えている統計量 $U_n:=(\underline{\theta},\bar{\theta})$ は，1 次の一般情報量損失は 0 になる，すなわち．$\Delta\to 0$ のとき

$$I_X^{(\alpha)}(\theta,\theta+\Delta)-I_{U_n}^{(\alpha)}(\theta,\theta+\Delta)=o(1)$$

となるが，U_n の 2 次の一般情報量損失は I_0 となることが分かる．このことは，2 次の次数まで情報量損失を無くすためには，さらに $f_0(x-\theta)$ の台の中央部分に関係する統計量 Z_1^* を付加する必要性を意味している．ここで，具体的に切断分布の場合に定理 4.2 を適用する(Akahira, 1996).

例 4.1　$X_1, X_2,\cdots, X_n,\cdots$ を互いに独立に，いずれも p.d.f.

$$f_0(x-\theta)=\begin{cases}ce^{-(x-\theta)^2/2} & (|x-\theta|<1),\\ 0 & (|x-\theta|\ge 1)\end{cases}$$

をもつ切断正規分布に従う確率変数列とする．ただし，$\theta\in\boldsymbol{R}^1$ で，c はある正の定数とする．このとき

$$\lim_{x\to -1+0}f_0(x)=\lim_{x\to 1-0}f_0(x)=ce^{-1/2},$$
$$h=\lim_{x\to 1-0}f_0'(x)=-\lim_{x\to -1+0}f_0'(x)=-ce^{-1/2}$$

になり，条件 (B1)〜(B4) が満たされる．よって，定理 4.2 から統計量 $(Z_1^*/(\sqrt{n}I_0),\underline{\theta},\bar{\theta})$ の 2 次の一般情報量損失は 0 になることがわかる．ここで

$$Z_1^* = \sqrt{n}(\bar{X} - \hat{\theta}^*), \quad \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i, \quad \hat{\theta}^* = \frac{1}{2}(\underline{\theta} + \bar{\theta}),$$

$$\underline{\theta} = \max_{1 \leq i \leq n} X_i - 1, \quad \bar{\theta} = \min_{1 \leq i \leq n} X_i + 1, \quad I_0 = 1 - 2ce^{-1/2}$$

になる.

例 4.2 $X_1, X_2, \cdots, X_n, \cdots$ を互いに独立に，いずれも p.d.f.

$$f_0(x-\theta) = \begin{cases} \dfrac{c}{1+(x-\theta)^2} & (|x-\theta| < 1), \\ 0 & (|x-\theta| \geq 1) \end{cases}$$

をもつ切断 Cauchy 分布に従う確率変数列とする．ただし，$\theta \in \boldsymbol{R}^1$ で，c はある正の定数とする．このとき

$$\lim_{x \to -1+0} f_0(x) = \lim_{x \to 1-0} f_0(x) = c/2,$$

$$h = \lim_{x \to 1-0} f_0'(x) = -\lim_{x \to -1+0} f_0'(x) = -c/2$$

となり，条件（B1)～(B4）が満たされる．よって，定理 4.2 から統計量 $(Z_1^*/(\sqrt{n}I_0), \underline{\theta}, \bar{\theta})$ の 2 次の一般情報量損失は 0 になることがわかる．ここで

$$Z_1^* = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \frac{X_i - \hat{\theta}^*}{1+(X_i - \hat{\theta}^*)^2}, \quad \hat{\theta}^* = \frac{1}{2}(\underline{\theta} + \bar{\theta}),$$

$$I_0 = 8c\int_0^1 \frac{x^2}{(1+x^2)^3} dx$$

となり，$\underline{\theta}, \bar{\theta}$ は例 4.1 と同じになる.

5. 全変動による情報量とその推定への応用

本節では，非正則な場合に適用可能な情報量として，全変動による情報量を考えて，推定量の漸近効率について論じる（Akahira, 2008).

まず，確率ベクトル $\boldsymbol{X} := (X_1, \cdots, X_n)$ の（σ-有限測度 μ_n に関する）j.p.d.f. を $f_X(\boldsymbol{x}, \theta)$ $(\theta \in \Theta)$ とし，Θ を $\boldsymbol{R}^1$ の開区間とするとき，$\theta, \theta + \Delta \in \Theta$ $(\Delta \neq 0)$ について

$$D_X(\theta, \theta+\Delta) := \frac{2}{|\Delta|} d_X(\theta, \theta+\Delta) \tag{5.1}$$

によって，$f_X(\cdot, \theta)$ に対する $f_X(\cdot, \theta+\Delta)$ の（識別）情報量を定義する．また，統計量 $T = T(\boldsymbol{X})$ の p.d.f. $f_T(t, \theta)$ によっても同様に $d_T(\theta, \theta+\Delta)$, $D_T(\theta, \theta+\Delta)$ を定義すれば，

$$D_T(\theta, \theta+\Delta) \leq D_X(\theta, \theta+\Delta)$$

が成り立つ．ここで,

$$e_T(\theta, \Delta) := \frac{D_T(\theta, \theta+\Delta)}{D_X(\theta, \theta+\Delta)} \tag{5.2}$$

を統計量 T の効率といい，$\Delta \to 0$ のとき $e_T(\theta, \Delta)$ の極限値を T の（1 次の）漸近効率という．いま，f_X に適当な正則条件を課せば，Fisher 情報量 $I_X(\theta) := E_\theta[\{(\partial/\partial\theta)\log f_X(\boldsymbol{X}, \theta)\}^2]$ につ

いて

$$\overline{\lim_{\Delta\to 0}} D_X^2(\theta, \theta+\Delta) \leq I_X(\theta)$$

になる．また，θ の一致推定量が存在するための必要条件として

$$\lim_{n\to\infty} |\Delta| D_X(\theta, \theta+\Delta) = 2$$

が得られる．いま，無限大に発散する正数の増加列を $\{c_n\}$ とする．このとき，任意の $\eta\in\Theta$ について，ある正数 δ が存在して

$$\lim_{L\to\infty} \overline{\lim_{n\to\infty}} \sup_{\theta:|\theta-\eta|<\delta} P_{\theta,n}\{c_n|\hat{\theta}_n(\boldsymbol{X})-\theta| \geq L\} = 0$$

となれば，$\hat{\theta}_n = \hat{\theta}_n(\boldsymbol{X})$ を θ の $\{c_n\}$・一致推定量という（竹内，1974；Akahira, 1975；Akahira and Takeuchi, 1981）．ただし，$dP_{\theta,n}/d\mu_n = f_X(\boldsymbol{x}, \theta)$ とする．このとき，$\{c_n\}$・一致推定量が一致推定量であることは明らか．さらに，θ の $\{c_n\}$・一致推定量が存在するための必要条件として次のことを得る．

定理 5.1　θ の $\{c_n\}$・一致推定量が存在すれば，任意の $\varepsilon>0$ と任意の $\theta\in\Theta$ に対して，$t(>0)$ が存在して

$$\underline{\lim_{n\to\infty}}\, c_n^{-1} D_X(\theta, \theta \pm t c_n^{-1}) \geq (2-\varepsilon)/t$$

が成り立つ．

定理 5.1 は Akahira（1975）の定理 3.3，Akahira（1995）の定理 2.1 の変形版になっている．

次に，特に，$X_1, X_2, \cdots, X_n, \cdots$ を互いに独立にいずれも（Lebesgue 測度に関する）p.d.f. $f_0(x-\theta)$ をもつ分布に従う確率変数列とする．ここで，第 4 節の条件（B1）の下では，$0<\Delta<b-a$ について

$$d_X(\theta, \theta-\Delta) \leq \left[1-\exp\left\{-\frac{n}{4} I_X^{(0)}(\theta, \theta-\Delta)\right\}\right]^{1/2}$$

となり，このことから θ の $\{c_n\}$・一致推定量が存在するための必要条件として次のことを得る（Akahira, 1995）．

定理 5.2　$X_1, X_2, \cdots, X_n, \cdots$ を互いに独立に，いずれも条件（B1）を満たす p.d.f. $f_0(x-\theta)$ をもつ分布に従う確率変数列とする．このとき，θ の $\{c_n\}$・一致推定量が存在すれば，任意の θ について，ある正数 t_0 が存在して，任意の $t(\geq t_0)$ に対して

$$\underline{\lim_{n\to\infty}}\, n I_X^{(0)}(\theta, \theta - t c_n^{-1}) > 0$$

である．

定理 5.1，5.2 における必要条件は，一致性の収束の最大次数 c_n^* を求めるとき，すなわち，c_n^* より速い（大きい）次数をもつ一致推定量が存在しないことを示すときに有用である（Akahira, 1975）．また，最近，Beckert and McFadden（2004）は，Hellinger 距離を用いて同様の結果を得ている．ここで，次の条件を仮定する．

（C1）　$0<\alpha\leq\beta<\infty$ について

α	c_n^*	$\{c_n^*\}$・一致推定量
$0<\alpha<2$	$n^{1/\alpha}$	$\{\min_{1\le i\le n} X_i+\max_{1\le i\le n} X_i-(a+b)\}/2$
$\alpha=2$	$\sqrt{n\log n}$	最尤推定量
$\alpha>2$	$\sqrt{n}$	最尤推定量

$$A' := \lim_{x\to a+0}(x-a)^{1-\alpha}f_0(x),\quad B' := \lim_{x\to b-0}(b-x)^{1-\beta}f_0(x)$$

が存在する．ただし，$0<A', B'<\infty$ とする．

(C2)　$0<\alpha\le\beta<\infty$ について

$$A'' := \lim_{x\to a+0}(x-a)^{2-\alpha}|f_0'(x)|,\quad B'' := \lim_{x\to b-0}(b-x)^{2-\beta}|f_0'(x)|$$

は有限確定である．また，$\alpha=2$ のとき $f_0''(x)$ は有界である．

このとき，定理5.2から一致性の収束の最大次数に関して次のことが成り立つ（Akahira, 1975, 1995; Akahira and Takeuchi, 1981).

定理 5.3　$X_1, X_2, \cdots, X_n, \cdots$ を互いに独立に，いずれも条件（B1),（B2),（C1),（C2）を満たすp.d.f. $f_0(x-\theta)$ をもつ分布に従う確率変数列とする．このとき，一致性の収束の最大次数 c_n^* および $\{c_n^*\}$・一致推定量は次のように与えられる．

例 5.1　$X_1, X_2, \cdots, X_n, \cdots$ を互いに独立に，いずれもp.d.f.

$$f_0(x-\theta)=\begin{cases} C_\alpha\{1-(x-\theta)^2\}^{\alpha-1} & (|x-\theta|<1), \\ 0 & (|x-\theta|\ge 1) \end{cases}$$

をもつ分布に従う確率変数列とする．ただし，$\alpha>0$, $C_\alpha := 1/\{2^{2\alpha-1}B(\alpha,\alpha)\}$ とする．このとき，$a=-1$, $b=1$, $\alpha=\beta$, $A'=B'=2^{\alpha-1}C_\alpha$, $A''=B''=2^{\alpha-1}C_\alpha|\alpha-1|$ とすれば条件（B1),（B2),（C1),（C2）は満たされるので，α の各値に対応して定理5.3の表のような最大次数 c_n^* をもつ $\{c_n^*\}$・一致推定量が存在する．

さらに，Kendall (1961), Nikulin (1993) によれば，例5.1において $\alpha=3/2$ としたときに，θ の推定問題を考えて，D. Bernoulli がいわゆる最尤法を用いて推定することを提案し，L. Euler はそれを認めた上で，それとは異なる推定量を提案した．それらは *Memoirs of the Academy of St. Petersburg, Acta Acad. Petrop.* (1777) にラテン語で掲載され，Kendall (1961) において英訳されている．この推定問題は興味深いが，なかなか難しい．一般に，条件（B1),（B2),（C1),（C2）において，$a=-1$, $b=1$, $1<\alpha=\beta<2$ とし，さらに

$$\lim_{x\to -1+0}(1+x)^{3-\alpha}f_0''(x),\quad \lim_{x\to 1-0}(1-x)^{3-\alpha}f_0''(x)$$

が有限確定であると仮定する．このとき，θ の漸近中央値不偏推定量 $\hat\theta_n=\hat\theta_n(\boldsymbol{X})$，すなわち θ について局所一様に

$$\lim_{n\to\infty} n^{1/\alpha}\left|P_{\theta,n}\{\hat\theta_n\le\theta\}-\frac{1}{2}\right|=0,\quad \lim_{n\to\infty} n^{1/\alpha}\left|P_{\theta,n}\{\hat\theta_n\ge\theta\}-\frac{1}{2}\right|=0$$

となる推定量 $\hat\theta_n$ の全体のクラスの中で，漸近中央値不偏になるように補正した最尤推定量は

漸近的有効にならないことが示されている (Akahira and Takeuchi, 1991b).

さて，ここで元に戻って，条件 (B1)，(B2) の下で考えると，まず，$f_X(\boldsymbol{x}, \theta)=\Pi_{i=1}^n f_0(x_i-\theta)$ となるから

$$2d_X(\theta, \theta+\Delta)=\int_{a+\Delta}^b\cdots\int_{a+\Delta}^b\left|\prod_{i=1}^n f_0(x_i-\Delta)-\prod_{i=1}^n f_0(x_i)\right|dx_1\cdots dx_n$$
$$+\left\{1-\left(\int_{a+\Delta}^b f_0(x)dx\right)^n\right\}+\left\{1-\left(\int_{a+\Delta}^b f_0(x-\Delta)dx\right)^n\right\} \quad (5.3)$$

になる．このとき，推定量の効率に関して，簡単な場合として次の例が挙げられる．

例 5.2 $X_1, X_2,\cdots, X_n,\cdots$ を互いに独立に，いずれも一様分布 $U(\theta-1/2, \theta+1/2)$ の p.d.f. をもつ分布に従う確率変数列とする．このとき (5.3) より，$d_X(\theta, \theta+\Delta)=1-(1-\Delta)^n$ になり，また θ の範囲の中央 $M:=(\min_{1\le i\le n} X_i+\max_{1\le i\le n} X_i)/2$ について，$U:=M-\theta$ の p.d.f. が

$$f_U(u)=\begin{cases} n(1-2|u|)^{n-1} & (|u|<1/2), \\ 0 & (|u|\ge 1/2) \end{cases}$$

となることから

$$d_M(\theta, \theta+\Delta)=1-(1-\Delta)^n$$

を得る．よって

$$\frac{D_M(\theta, \theta+\Delta)}{D_X(\theta, \theta+\Delta)}=\frac{d_M(\theta, \theta+\Delta)}{d_X(\theta, \theta+\Delta)}=1$$

になり M の効率は 1 である．

もっと一般の場合に統計量の漸近効率を求めるために，次のことが有用である．

定理 5.4 条件 (B1)，(B2) を仮定する．また $\Delta=O(1/n)$ で，$\Delta>0$ とすれば，$n\to\infty$ のとき

$$\frac{1}{n}D_X(\theta, \theta+\Delta)=\frac{1}{n\Delta}|e^{-n\Delta\{f_0(b-0)-f_0(a+0)\}}-1|\left(\int_{a+\Delta}^b f_0(x)dx\right)^n$$
$$+\frac{1}{n\Delta}\left\{1-\left(\int_{a+\Delta}^b f_0(x)dx\right)^n\right\}$$
$$+\frac{1}{n\Delta}\left\{1-\left(\int_{a+\Delta}^b f_0(x-\Delta)dx\right)^n\right\}+o(1) \quad (5.4)$$

が成り立つ．

証明については，(5.3) を Taylor 展開すれば (5.4) を得る．

例 5.3 $X_1, X_2,\cdots, X_n,\cdots$ を互いに独立に，いずれも p.d.f.

$$f_0(x-\theta)=\begin{cases} ce^{-(x-\theta)} & (|x-\theta|<1/2), \\ 0 & (|x-\theta|\ge 1/2) \end{cases}$$

をもつ分布に従う確率変数列とする．ただし，$c=1/(e^{1/2}-e^{-1/2})$ とする．このとき，

$$f_0\left(\frac{1}{2}-0\right)=ce^{-1/2},\quad f_0\left(-\frac{1}{2}+0\right)=ce^{1/2}$$

になり，$\Delta=t/n(t>0)$ とすると

$$e^{-n\Delta\{f_0(1/2-0)-f_0(-1/2+0)\}}=e^{-tc\{e^{-1/2}-e^{1/2}\}}=e^t$$

となる．また，$n\to\infty$のとき

$$\left\{\int_{-\frac{1}{2}+\Delta}^{\frac{1}{2}} f_0(x)\,dx\right\}^n=\left(1-\int_{-\frac{1}{2}}^{-\frac{1}{2}+\frac{t}{n}} ce^{-x}dx\right)^n=e^{-ce^{1/2}t}+o(1),$$

$$\left\{\int_{-\frac{1}{2}+\Delta}^{\frac{1}{2}} f_0(x-\Delta)\,dx\right\}^n=\left(1-\int_{\frac{1}{2}}^{\frac{1}{2}+\frac{t}{n}} ce^{-(x-\Delta)}dx\right)^n=e^{-ce^{-1/2}t}+o(1)$$

になる．ここで$\alpha:=ce^{1/2}$とおくと$ce^{-1/2}=\alpha-1$となるから，(5.4) より$n\to\infty$のとき

$$\frac{1}{n}D_X\left(\theta,\theta+\frac{t}{n}\right)=\frac{2}{t}(1-e^{-\alpha t})+o(1) \tag{5.5}$$

となる．一方，θの最尤推定量$\hat{\theta}_{ML}:=\min_{1\le i\le n}X_i+(1/2)$について，$U=\hat{\theta}_{ML}-\theta$の p.d.f. は

$$f_U(u)=\begin{cases} nc^n e^{-u+(1/2)}(e^{-u+(1/2)}-e^{-1/2})^{n-1} & (0<u<1),\\ 0 & (\text{その他})\end{cases}$$

となるから，$n\to\infty$のとき

$$2d_{\hat{\theta}_{ML}}\left(\theta,\theta+\frac{t}{n}\right)=\int_0^1\left|f_U\left(u-\frac{t}{n}\right)-f_U(u)\right|du=2(1-e^{-\alpha t})+o(1)$$

となる．よって，(5.1)，(5.5) より$n\to\infty$のとき

$$\frac{D_{\hat{\theta}_{ML}}\left(\theta,\theta+\frac{t}{n}\right)}{D_X\left(\theta,\theta+\frac{t}{n}\right)}=1+o(1)$$

となり，$\hat{\theta}_{ML}$の漸近効率は1になる．

上記において，1次の漸近効率について述べたが，(5.2) において

$$e_T(\theta,\Delta)=1-\gamma\Delta+o(\Delta)\quad(\Delta\to 0)$$

となるとき，$1-\gamma\Delta$をTの2次の漸近効率と定義でき，さらに高次まで考えることもできる (Akahira, 2008).

6. おわりに

本稿において論じた情報量は，非正則な場合を含む推定論の広い範囲を対象にできるようなものである．実際には正則な場合の推定論は，すでにかなり確立されているが，非正則な場合は，まだ十分とは言い難い面もある．たとえば，Chapman-Robbins 型の不等式についても適切な情報量を導入することによってもっと見通し良く論じることができると思われるが，その下界の達成についてはやや面倒な議論になるであろう．また，Bhattacharyya 型の不等式についても同様に議論を進めることもできる．そして，推定量の Bayes リスクに関する情報不等式を導出するために有用な情報量を考える必要がある (Vincze, 1992; Akahira and Ohyauchi, 2003; Ohyauchi, 2004)．さらに，最近，非正則な場合の逐次推定についても論じられている (Akahira and Koike, 2005; Koike, 2007)．その他にも，非正則な場合の特殊な場合と見なされる特異モデルの尤度に基づく議論が行われている (福水・栗木 他, 2004)．ともあれ，適切な情報量に

よる包括的な議論は推定論の構造そのものを明確にしてくれるであろう.

参考文献

Akahira, M. (1975). Asymptotic theory for estimation of location in non-regular cases, I: Order of convergence of consistent estimators. *Rep. Stat. Appl. Res., JUSE* **22** (3), 3–19.

Akahira, M. (1986). *The Structure of Asymptotic Deficiency of Estimators*. Queen's Papers in Pure and Appl. Math., **75**, Queen's University Press, Kingston, Canada.

Akahira, M. (1993). Asymptotics on the statistics for a family of non-regular distributions, *Statistical Sciences and Data Analysis* (K. Matusita et al. Eds), VSP Internat. Sci. Publ., Zeist (Netherlands), 357–364.

Akahira, M. (1995). The amount of information and the bound for the order of consistency for a location parameter family of densities, *Proceedings of the 2nd Gauss Symposium. Conference B: Statistical Sciences*, Sympos. Gaussiana, de Gruyter, Berlin, 303–311.

Akahira, M. (1996). Loss of information of a statistic for a family of non-regular distributions, *Ann. Inst. Statist. Math.*, **48**, 349–364.

赤平昌文 (2006).「統計的推定の高次漸近理論の構造」『数学』**58**(1), 岩波書店, 1-20.

Akahira, M. (2008). A definition of information amount and its application to non-regular estimation. In preparation.

Akahira, M. and Koike, K. (2005). Sequential interval estimation of a location parameter with the fixed width in the uniform distribution with unknown scale parameter, *Sequential Analysis* **24**(1), 63–75.

Akahira, M. and Ohyauchi, N. (2003). An information inequality for the Bayes risk applicable to non-regular cases, *Proc. Sympos., Res. Inst. Math. Sci.*, **1334**, Kyoto University, 183–191.

Akahira, M. and Takeuchi, K. (1981). *Asymptotic Efficiency of Statistical Estimators: Concepts and Higher Order Asymptotic Efficiency*. Lecture Notes in Statistics **7**, Springer, New York.

Akahira, M. and Takeuchi, K. (1985). A note on the minimum variance unbiased estimation when the Fisher information is infinity, *Rep. Stat. Appl. Res., JUSE*, **32**(3), 17–22. Also included In: "*Joint Statistical Papers of Akahira and Takeuchi*," World Scientific, New Jersey, 2003, 231–236.

Akahira, M. and Takeuchi, K. (1991a). A definition of information amount applicable to nonregular cases, *Journal of Computing and Information* **2**(1), 71–92. Also included In: "*Joint Statistical Papers of Akahira and Takeuchi*," World Scientific, New Jersey, 2003, 455–476.

Akahira, M. and Takeuchi, K. (1991b). Asymptotic efficiency of estimators for a location parameter family of densities with the bounded support, *Rep. Stat. Appl. Res., JUSE*, **38**(1), 1–9. Also included In: "*Joint Statistical Papers of Akahira and Takeuchi*," World Scientific, New Jersey, 2003, 446–454.

Akahira, M. and Takeuchi, K. (1995). *Non-Regular Statistical Estimation*. Lecture Notes in Statistics **107**, Springer, New York.

安藤洋美 (1985).「イエルジイ・ネイマンの生涯②」『Basic』数学8月号, 現代数学社, 77–81.

Beckert, W. and McFadden, D. (2004). Maximal uniform convergence rates in parametric estimation problems, *Birkbeck Working Papers in Economics & Finance* (*BWPEF*), 0405, University of London.

Fisher, R. A. (1925). Theory of statistical estimation, *Proc. Camb. Phil. Soc.*, **22**, 700–725.

Fisher, R. A. (1934). Two new properties of mathmatical likelihood, *Proc. Roy. Soc.*, A, **144**, 285–307.

福水健次, 栗木哲, 竹内啓, 赤平昌文 (2004).『特異モデルの統計学』岩波書店.

Ibragimov, I. A. and Has'minskii, R. Z. (1981). *Statistical Estimation*. Springer, New York.

Kendall, M. G. (1961). Studies in the history of probability and statistics XI. Daniel Bernoulli on maximum likelihood, *Biometrika*, **48**(1/2), 1-18.

Koike, K. (2007). Sequential point estimation of the location parameter in the location-scale family of non-regular distributions, *Sequential Analysis*, **26**(4), 383–393.

Kullback, S. (1959). *Information Theory and Statistics*. Wiley, New York.

Le Cam, L. (1990). On the standard asymptotic confidence ellipsoids of Wald, *Int. Statist. Rev.*, **58**(2), 129–152.

Matusita, K. (1955). Decision rules based on the distance for problem of fit, two samples and estimation, *Ann. Math. Statist.*, **26**, 631–640.

Nikulin, M. S. (1993). About one problem of D. Bernoulli and L. Euler from the theory of point estimation. Unpublished manuscript.

Ohyauchi, N. (2004). The Vincze inequality for the Bayes risk, *J. Japan Statist. Soc.*, **34**(1), 65–74.

大谷内奈穂, 赤平昌文 (2007). Maximal orders of convergence of consistency in terms of measures of diversity,

京都大学『数理解析研究所講究録』**1560**, 38-46.

Pardo, L. (2006). *Statistical Inference Based on Divergence Measures*. Chapman & Hall/CRC, Boca Raton.

Rényi, A. (1961). On measures of entropy and information, *Proc. of the Fourth Berkeley Symp. on Math. Statist. and Prob.*, **1**, 547-561.

Strasser, H. (1985). *Mathematical Theory of Statistics*. Walter de Gruyter, New York.

竹内啓 (1974). 『統計的推定の漸近理論』教育出版.

Vincze, I. (1992). On nonparametric Cramér-Rao inequalities, in: *Order Statistics and Nonparametrics: Theory and Application*, (P. K. Sen and I. A. Salama, eds.), Elsevier Science Publishers B. V., 439-454.