数理解析研究所講究録 1439 巻 2005 年 247-253

Lower bounds for the Bayes risk of unbiased estimators in non-regular cases

筑波大・数理物質 大谷内奈穂 (Nao Ohyauchi)
筑波大・数理物質 赤平 昌文 (Masafumi Akahira)

1 はじめに

統計的推定論において, 適当な正則条件の下では不偏推定量の分散の限界を与える情報不等式として, Cramér-Rao (C-R), Bhattacharyya の不等式等が知られている. また正則条件が必ずしも成り立たないような非正則の場合には, 位置母数をもつ切断分布族に適用可能な Hammersley-Chapman-Robbins の不等式がよく知られている (Hammersley(1950), Chapman and Robbins(1951)). また, 分布の密度の台が一方向に動く母数をもつ一方向型分布族において, 任意の不偏推定量の分散の下限は, 母数の任意に固定された値で 0 になることなどが示されている (Akahira and Takeuchi(1995)).

一般に, 母数空間の任意の 2 点における任意の不偏推定量の分散の凸結合に対する下界は, Vincze(1992) によって, C-R の不等式に基づいて間接的に求められた. 最近, その下界は Lagrange 法を用いて直接的に導出された (Akahira and Ohyauchi(2003), Ohyauchi(2004)). 本論では, もっと一般に, 任意の不偏推定量の混合事前確率測度に関する Bayes リスクに対する情報不等式を Vincze(1992) と類似の方法で導出し, その特別な場合として Vincze の不等式を導く. また, 不偏推定量の Bayes リスクに対する下界も求め, さらに, 非正則な場合の適用例も挙げる.

2 Bayes リスクに関する情報不等式

確率ベクトル $\boldsymbol{X} := (X_1, \cdots, X_n)$ の (σ-有限測度 μ に関する) 同時密度を $f_{\boldsymbol{X}}(\boldsymbol{x}, \theta)$ ($\theta \in \Theta$) とする. ただし, $\theta \in \Theta \subset \mathbf{R}^1$ とし, $\boldsymbol{x} := (x_1, \cdots, x_n)$ とする. いま, Π_1 と Π_2 を Θ 上の事前確率測度とし, $\Pi_{12} := p\Pi_1 + q\Pi_2$ $(0 < p < 1,\ \ q = 1 - p)$ を考える. また, θ の実数値関数 $g(\theta)$ の推定量 $\hat{g}(\boldsymbol{X})$ の 2 乗損失によるリスクを

$$R(\theta, \hat{g}) := E_\theta\left[\{\hat{g}(\boldsymbol{X}) - g(\theta)\}^2\right] = \int_{\mathcal{X}} \{\hat{g}(\boldsymbol{x}) - g(\theta)\}^2 f_{\mathbf{X}}(\mathbf{x}, \theta) d\mu(\mathbf{x})$$

とする. ただし, $\mathcal{X}$ は $\mathbf{X}$ の標本空間とする. このとき, 事前確率測度 Π_{12} に関する推定量 $\hat{g}(\boldsymbol{X})$ の Bayes リスクは

$$r(\Pi_{12}, \hat{g}) := \int_\Theta R(\theta, \hat{g}) d\Pi_{12}(\theta)$$

になる. また,

$$\int_\Theta g(\theta) d\Pi_i(\theta) = g_i \ \ (i = 1, 2)$$

で $g_1 \neq g_2$ とする. ここで, $\hat{g}$ を $g(\theta)$ の不偏推定量とし,

$$f_p(\mathbf{x}) := \int_\Theta f_{\mathbf{X}}(\mathbf{x}, \theta) d\Pi_{12}(\theta) = p \int_\Theta f_{\mathbf{X}}(\mathbf{x}, \theta) d\Pi_1(\theta) + q \int_\Theta f_{\mathbf{X}}(\mathbf{x}, \theta) d\Pi_2(\theta)$$

とする. そして, p の推定量として,

$$\hat{p}(\mathbf{X}) := \{\hat{g}(\mathbf{X}) - g_2\}/(g_1 - g_2)$$

を考える. このとき, $\hat{p}$ は密度 f_p の下で p の不偏推定量, すなわち, 任意の p $(0 < p < 1)$ について $E_p(\hat{p}) = p$ となり, また, f_p の下での $\hat{p}$ の分散は

$$V_p(\hat{p}) = V_p(\hat{g})/(g_1 - g_2)^2 \tag{2.1}$$

になる. 一方, Cramér-Rao の不等式から

$$V_p(\hat{p}) \geq 1/E_p\left[\{(\partial/\partial p)\log f_p(\mathbf{X})\}^2\right] =: 1/I_p \tag{2.2}$$

となる. ただし,

$$I_p = \int_{\mathcal{X}} \frac{\{\int_\Theta f_{\mathbf{X}}(\mathbf{x},\theta)d\Pi_1(\theta) - \int_\Theta f_{\mathbf{X}}(\mathbf{x},\theta)d\Pi_2(\theta)\}^2}{p\int_\Theta f_{\mathbf{X}}(\mathbf{x},\theta)d\Pi_1(\theta) + q\int_\Theta f_{\mathbf{X}}(\mathbf{x},\theta)d\Pi_2(\theta)} d\mu(\mathbf{x})$$

とする. このとき, 次の定理が成立する.

定理 $g(\theta)$ の任意の不偏推定量 $\hat{g}$ の Bayes リスクに関する情報不等式

$$r(\Pi_{12},\hat{g}) \geq (g_1 - g_2)^2\left(\frac{1}{I_p} - pq\right) - (p\sigma_1^2 + q\sigma_2^2) =: B_{12}(p) \tag{2.3}$$

が成立する. ただし, 各 $i = 1, 2$ について σ_i^2 は $g(\theta)$ の事前分散, つまり

$$\sigma_i^2 := \int_\Theta \{g(\theta) - g_i\}^2 d\Pi_i(\theta)$$

とする.

証明の概略 各 $i = 1, 2$ に対して

$$V_i := \int_\Theta V_\theta(\hat{g})d\Pi_i(\theta)$$

とおくと, 密度 f_p の下で $\hat{g}$ の分散は

$$V_p(\hat{g}) = pV_1 + qV_2 + p\sigma_1^2 + q\sigma_2^2 + pq(g_1 - g_2)^2$$

となるので, (2.1), (2.2) から

$$pV_1 + qV_2 \geq \frac{(g_1 - g_2)^2}{I_p} - (p\sigma_1^2 + q\sigma_2^2) - pq(g_1 - g_2)^2 \tag{2.4}$$

となる. よって, Π_{12} に関する $\hat{g}$ の Bayes リスクは

$$r(\Pi_{12},\hat{g}) = \int_\Theta V_\theta(\hat{g})d\Pi_{12}(\theta) = pV_1 + qV_2$$

となり, (2.4) から

$$r(\Pi_{12},\hat{g}) \geq (g_1-g_2)^2\left(\frac{1}{I_p}-pq\right)-(p\sigma_1^2+q\sigma_2^2)$$

を得る. □

系 Bayes リスクに対する下界 $B_{12}(p)$ について, 次の不等式

$$\sup_{\Pi_1,\Pi_2:\int_\Theta g(\theta)d\Pi_i(\theta)=g_i\ (i=1,2)} B_{12}(p) \geq B_{12}^*(p) := (g_1-g_2)^2\left(\frac{1}{I_p}-pq\right) \geq B_{12}(p)$$

が成立する.

証明については, 各 $i=1,2$ について事前確率測度 Π_i を $\Pi_i(\{\theta|g(\theta)=g_i\})=1$ となるようにとれば, $B_{12}(p)=B_{12}^*(p)$ となるから, (2.3) よりその不等式が成り立つ.

3 特別な場合

いま, 特に, θ_1, $\theta_2\in\Theta$ で $\theta_1\neq\theta_2$ とし, 各 $i=1,2$ について事前確率測度 Π_i を $\Pi_i(\{\theta_i\})=1$ となるようにとる. このとき,

$$I_{\boldsymbol{X}}^{(p)}(\theta_1,\theta_2) := \int_{\mathcal{X}}\frac{\{f_{\boldsymbol{X}}(\boldsymbol{x},\theta_1)-f_{\boldsymbol{X}}(\boldsymbol{x},\theta_2)\}^2}{pf_{\boldsymbol{X}}(\boldsymbol{x},\theta_1)+qf_{\boldsymbol{X}}(\boldsymbol{x},\theta_2)}d\mu(\boldsymbol{x})$$

とすると, 定理と系から, $g(\theta)$ の任意の不偏推定量 $\hat{g}$ に対して

$$r(\Pi_{12},\hat{g}) \geq \{g(\theta_1)-g(\theta_2)\}^2\left(\frac{1}{I_{\boldsymbol{X}}^{(p)}(\theta_1,\theta_2)}-pq\right) \tag{3.1}$$

が成立する (Vincze(1992)).

また, 条件

(A) $$E_{\theta_i}(\hat{g})=g(\theta_i)\quad(i=1,2)$$

の下で, Π_{12} に関する Bayes リスクを最小にする推定量 $\hat{g}=\hat{g}^*(\boldsymbol{X})$ を求める問題を考える. まず, 条件 (A) から

$$\int_{\mathcal{X}}\hat{g}(\boldsymbol{x})\{pf_{\boldsymbol{X}}(\boldsymbol{x},\theta_1)+qf_{\boldsymbol{X}}(\boldsymbol{x},\theta_2)\}d\mu(\mathbf{x})=pg(\theta_1)+qg(\theta_2)=:\eta \tag{3.2}$$

になる. 次に

$$h^{(p)}(\boldsymbol{x};\theta_1,\theta_2) := p\{f_{\boldsymbol{X}}(\boldsymbol{x},\theta_1)-f_{\boldsymbol{X}}(\boldsymbol{x},\theta_2)\}$$

とおくと,

$$\int_{\mathcal{X}}\hat{g}(\boldsymbol{x})h^{(p)}(\boldsymbol{x};\theta_1,\theta_2)d\mu(\boldsymbol{x})=p\{g(\theta_1)-g(\theta_2)\}=:\delta \tag{3.3}$$

を得る. ここで, (3.2), (3.3) は条件 (A) と同値であることに注意. このとき, (3.2), (3.3) の下で, Π_{12} に関する Bayes リスクを最小にする推定量 $\hat{g}^*$ を Lagrange 法を用いて考えれば, 条件 (A) をみたす $g(\theta)$ の任意の推定量 $\hat{g}$ に対して, Bayes リスクに関する情報不等式 (3.1) が成立し, その下界が $g(\theta)$ の推定量

$$\hat{g}^*(\boldsymbol{X}) := \eta + \frac{\{g(\theta_1) - g(\theta_2)\}\{f_{\boldsymbol{X}}(\boldsymbol{X}, \theta_1) - f_{\boldsymbol{X}}(\boldsymbol{X}, \theta_2)\}}{I_{\boldsymbol{X}}^{(p)}(\theta_1, \theta_2)\{pf_{\boldsymbol{X}}(\boldsymbol{X}, \theta_1) + qf_{\boldsymbol{X}}(\boldsymbol{X}, \theta_2)\}}. \tag{3.4}$$

によって達成されることも分かる (Akahira and Ohyauchi(2003), Ohyauchi(2004)).

例 1 $X_1, \cdots, X_n$ をたがいに独立に, いずれも一様分布 $U(\theta - (1/2), \theta + (1/2))$ に従う確率変数とする. いま, $g(\theta) = \theta$, $\theta_1 < \theta_2 < \theta_1 + 1$ とする. このとき

$$I_{\boldsymbol{X}}^{(p)}(\theta_1, \theta_2) = \frac{1}{pq}\{1 - (1 - (\theta_2 - \theta_1))^n\}$$

より, (3.1) の下界は

$$pq(\theta_1 - \theta_2)^2 \left\{\frac{1}{1 - (1 - (\theta_2 - \theta_1))^n} - 1\right\} \tag{3.5}$$

で与えられる. ここで, $j = 1, 2$ に対して

$$S_j := \{\boldsymbol{x} : \max_{1 \le i \le n} x_i - \frac{1}{2} \le \theta_j \le \min_{1 \le i \le n} x_i + \frac{1}{2}\}$$

とすると, (3.4) より下界 (3.5) は, θ の推定量

$$\hat{\theta}^*(\boldsymbol{X}) := p\theta_1 + q\theta_2 + \frac{pq(\theta_1 - \theta_2)}{1 - (1 - (\theta_2 - \theta_1))^n} \left\{\frac{1}{p}\chi_{S_1 - (S_1 \cap S_2)}(\boldsymbol{X}) + \frac{1}{q}\chi_{S_2 - (S_1 \cap S_2)}(\boldsymbol{X})\right\}$$

によって達成される. ただし, $\chi_A(x)$ は $A(\subset \mathcal{X})$ の定義関数とする.

4 不偏推定量の Bayes リスクに対する下界

前節において, 情報不等式 (3.1) の適用例として一様分布の場合を挙げたが, 少し複雑な分布になると一般の n については $I_{\boldsymbol{X}}^{(p)}$ の計算が難しくなることが多い. そこで, 本節では計算し易い (Bayes リスクに対する) 下界を導出し, 適用例も挙げる.

まず, $X_1, \cdots, X_n$ は, たがいに独立にいずれも (Lebesgue 測度に関する) 密度 $p(x, \theta)$ をもつ分布からの実確率変数とする. ただし, $\theta \in \Theta \subset \mathbf{R}^1$ とし, $A(\theta) := \{x \ : \ p(x, \theta) > 0\}$ とする. ここで, $\theta, \theta + \delta \in \Theta$ に対して, 情報量として

$$I_\delta(\theta) := \int_{A(\theta) \cup A(\theta + \delta)} \frac{\{p(x, \theta + \delta) - p(x, \theta)\}^2}{2\{p(x, \theta) + p(x, \theta + \delta)\}} dx$$

と定義すると,

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left| \prod_{i=1}^n p(x_i, \theta + \delta) - \prod_{i=1}^n p(x_i, \theta) \right| dx_1 \cdots dx_n \le 2\{(1 + I_\delta(\theta))^n - 1\}^{1/2} \tag{4.1}$$

となる (Akahira(1975)). ここで, (3.1) において, $p = q = 1/2$, $g(\theta) = \theta$, $\theta_1 = \theta + \delta$, $\theta_2 = \theta$ とする. このとき, $I_X^{(p)}(\cdot,\cdot)$ の定義と (4.1) より

$$
\begin{aligned}
I_{\mathbf{X}}^{(1/2)}(\theta+\delta,\theta) &= \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\frac{2\{\prod_{i=1}^n p(x_i,\theta+\delta)-\prod_{i=1}^n p(x_i,\theta)\}^2}{\prod_{i=1}^n p(x_i,\theta+\delta)+\prod_{i=1}^n p(x_i,\theta)}dx_1\cdots dx_n \\
&\leq 2\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\left|\prod_{i=1}^n p(x_i,\theta+\delta)-\prod_{i=1}^n p(x_i,\theta)\right|dx_1\cdots dx_n \\
&\leq 4\left\{(1+I_\delta(\theta))^n-1\right\}^{1/2}
\end{aligned}
$$

になる. よって, (3.1) より, θ の任意の不偏推定量 $\hat{\theta}$ に対して

$$
r\left(\Pi_{12},\hat{\theta}\right) \geq \frac{\delta^2}{4}\left[\frac{1}{\{(1+I_\delta(\theta))^n-1\}^{1/2}}-1\right] \tag{4.2}
$$

が成り立つ.

例 2 $X_1,\cdots,X_n$ をたがいに独立に, いずれも密度

$$
p(x,\theta)=\begin{cases} r & (0\leq x\leq\theta), \\ s:=1-r & (\theta<x<\theta+1), \\ r & (\theta+1\leq x\leq 2), \\ 0 & (\text{その他}) \end{cases}
$$

に従う確率変数とする. ただし, $0<\theta<1$, $0<r<s<1$ とする. このとき, $\delta>0$ について

$$
p(x,\theta+\delta)-p(x,\theta)=\begin{cases} s-r & (\theta<x<\theta+\delta), \\ r-s & (\theta+1<x<\theta+\delta+1), \\ 0 & (\text{その他}) \end{cases}
$$

となるから

$$
I_\delta(\theta)=\frac{1}{2}\left\{\int_\theta^{\theta+\delta}(s-r)^2dx+\int_{\theta+1}^{\theta+1+\delta}(r-s)^2dx\right\}=(s-r)^2\delta \tag{4.3}
$$

になる. よって, (4.2), (4.3) より, θ の任意の不偏推定量 $\hat{\theta}$ について

$$
r(\Pi_{12},\hat{\theta}) \geq \frac{\delta^2}{4}\left[\frac{1}{\{(1+(s-r)^2\delta)^n-1\}^{1/2}}-1\right] =: B^{(1/2)}(\theta+\delta,\theta)
$$

となる. いま, $\delta=t/n(>0)$ とおけば, 十分大きな n について

$$
B^{(1/2)}\left(\theta+\frac{t}{n},\theta\right) \approx \frac{t^2}{4n^2}\left[\frac{1}{\{e^{(s-r)^2t}-1\}^{1/2}}-1\right] \tag{4.4}
$$

になる．そこで，特に $r=1/3$, $s=2/3$ とすれば，(4.4) から $n^2B^{(1/2)}(\theta+t/n,\theta)$ は漸近的に最大値 1.38018 をもつことが分かる．

例 3 $X_1, X_2, \cdots, X_n, \cdots$ をたがいに独立に，いずれも密度

$$p(x,\theta)=\begin{cases} ce^{-(x-\theta)} & (\theta<x<\theta+1),\\ 0 & (\text{その他}) \end{cases}$$

をもつ切断指数分布に従う確率変数列とする．ただし，$c=e/(e-1)$, $\theta\in\mathbf{R}^1$ とする．このとき，$0<\delta<1$ について

$$\begin{aligned} I_\delta(\theta) &= \int_\theta^{\theta+\delta+1}\frac{\{p(x,\theta+\delta)-p(x,\theta)\}^2}{2\{p(x,\theta)+p(x,\theta+\delta)\}}dx \\ &= \frac{c}{2}(e^\delta-1)\left\{\frac{1}{e^\delta}+\frac{1}{e}+\frac{e^\delta-1}{e^\delta+1}\left(\frac{1}{e^\delta}-\frac{1}{e}\right)\right\} \\ &= \frac{e+1}{2(e-1)}\delta+O\left(\delta^2\right)=: k\delta+O\left(\delta^2\right) \end{aligned}$$

になる．ただし，$k=(e+1)/\{2(e-1)\}(\approx 1.082)$ とする．よって，(4.2), (4.3) より θ の任意の不偏推定量 $\hat\theta$ について

$$r\left(\Pi_{12},\hat\theta\right)\geq\frac{\delta^2}{4}\left[\frac{1}{\{(1+k\delta+O\left(\delta^2\right))^n-1\}^{1/2}}-1\right]=: B^{(1/2)}(\theta+\delta,\theta)$$

となる．いま，$\delta=t/n(>0)$ とおくと，十分大きい n について

$$B^{(1/2)}\left(\theta+\frac{t}{n},\theta\right)\approx\frac{t^2}{4n^2}\left\{\frac{1}{(e^{kt}-1)^{1/2}}-1\right\} \tag{4.5}$$

になる．このとき，(4.5) から $n^2B^{(1/2)}(\theta+t/n,\theta)$ は漸近的に最大値として約 0.01455 をもつことが分かる．

5 おわりに

本論において，例 1，例 2 のような非正則な場合に適用可能な情報不等式について論じた．しかし，一般には不偏推定量の Bayes リスクに関する情報不等式による下界は達成されない．今後，さらに適切な情報量を導入して情報不等式を改良する余地はある．

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**, 8–26.

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. (1995). *Non-Regular Statistical Estimation.* Lecture Notes in Statistics 107, Springer, New York.

Chapman, D. G. and Robbins, H. (1951). Minimum variance estimation without regularity assumptions. *Ann. Math. Statist.*, **22**, 581–586.

Hammersley, J. M. (1950). On estimating restricted parameters. *J. Roy. Statist. Soc., Ser. B*, **12**, 192–240.

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

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