数理解析研究所講究録
第623巻 1987年 43-53

Amount of Information in Non-Regular Estimation

赤平 昌文 (Masafumi Akahira)

はじめに.

標本および統計量に含まれる情報量という概念は、統計的推測の理論において重要な役割を果している (e.g. Fisher, 1925, 1934; Kullback, 1959). しかし従来の議論は、ほとんどすべて正則な場合に限られていたように思われる. ここでは非正則な場合を十分考慮に入れて、情報量の概念とその適用について論じてみたい.

1. 情報量

Xを確率変数とし、$(\mathcal{X}, \mathcal{B})$を標本空間とする. また$P, Q$を$X$の確率分布とし、それらはある$\sigma$-有限測度$\mu$に関して絶対連続であると仮定する. このとき$P, Q$の間の情報量を

$$I_X(P, Q) = -8 \log \int_{\mathcal{X}} \left(\frac{dP}{d\mu} \cdot \frac{dQ}{d\mu} \right)^{1/2} d\mu$$

によって定義する（竹内，1983）．ここで右辺における積分値は類似度（affinity）と呼ばれている（Matusita, 1955）．上の情報量は、測度μのとり方には依存しない．実際νがμに関して絶対連続ならば、

$$\int_{\mathcal{X}}\left(\frac{dP}{d\nu}\cdot\frac{dQ}{d\nu}\right)^{1/2}d\nu=\int_{\mathcal{X}}\left(\frac{dP}{d\mu}\cdot\frac{dQ}{d\mu}\right)^{1/2}d\mu$$

になる．またμとして$P+Q$をとれば、$I_X(P,Q)$は一意的に定義される．さらにPとQが同値でないならば、$I_X(P,Q)>0$である．P,Qがdisjointでないこと、すなわち任意の$\mathcal{B}$-可測集合Bに対して$P(B)Q(B)+(1-P(B))(1-Q(B))>0$であることは、$I_X(P,Q)<\infty$であることと同値である．このことは、$\mu(\{x: dP/d\mu\,(x)>0\}\cap\{x: dQ/d\mu\,(x)>0\})>0$と同値になることから示される．

次に、X, Yをたがいに独立な確率変数とするとき、(P_1, Q_1)，(P_2, Q_2)を(X, Y)の確率分布とすれば、

$$I_{X,Y}(P_1\times Q_1,\ P_2\times Q_2)=I_X(P_1, P_2)+I_Y(Q_1, Q_2)$$

が成り立つ．ただし各$i=1,2$について、$P_i\times Q_i$はP_iとQ_iの直積測度とする．

また$T=t(X)$を統計量とするとき、確率分布P, QからTによって誘導された分布をP_T，Q_Tとし、$I_T(P_T, Q_T)$を単に$I_T(P,Q)$で表わす．このとき次のことが成り立つ．

定理. $I_T(P,Q) \le I_X(P,Q)$ が成り立つ．ここで等式が成り立つのは、Tが P, Q に対して十分であるときに限る．

次に上で定義された情報量 $I_X(P,Q)$ の中の定数8が、Fisher情報量との関係から出てくることを示す．分布が実母数 θ に依存するとき、$I_X(P_{\theta_1}, P_{\theta_2})$ を単に $I_X(\theta_1, \theta_2)$ で表わす．ある母数 θ_0 の近傍において、P_θ は P_{θ_0} に関して絶対連続でほとんどすべての x について $dP_\theta/dP_{\theta_0}(x)$ は θ に関して連続微分可能であると仮定する．このとき μ として P_{θ_0} をとれば十分小さな $\Delta > 0$ に対して、次のことが成り立つ．

$$
\begin{aligned}
I(\theta_0, \theta_0+\Delta) &= -8 \log \left[\int_{\mathfrak{X}} \left(\frac{dP_{\theta_0+\Delta}}{dP_{\theta_0}} \right)^{1/2} dP_{\theta_0} \right] \\
&= -8 \log \left[\int_{\mathfrak{X}} \left\{ 1 + \frac{d(P_{\theta_0+\Delta} - P_{\theta_0})}{dP_{\theta_0}} \right\}^{1/2} dP_{\theta_0} \right] \\
&= -8 \log \left[1 - \frac{1}{8} \int_{\mathfrak{X}} \left\{ \frac{d(P_{\theta_0+\Delta} - P_{\theta_0})}{dP_{\theta_0}} \right\}^2 dP_{\theta_0} + o(\Delta^2) \right] \\
&= \int_{\mathfrak{X}} \left\{ \left[\frac{\partial}{\partial\theta} \left(\frac{dP_\theta}{dP_{\theta_0}} \right) \right]_{\theta=\theta_0} \right\}^2 dP_{\theta_0} \cdot \Delta^2 + o(\Delta^2) \\
&= E_{\theta_0} \left[\left\{ \left[\frac{\partial}{\partial\theta} \log \frac{dP_\theta}{dP_{\theta_0}} \right]_{\theta=\theta_0} \right\}^2 \right] \Delta^2 + o(\Delta^2) \\
&= I(\theta_0)\Delta^2 + o(\Delta^2)
\end{aligned}
$$

ただし、$I(\theta_0)$ は Fisher 情報量である．

2. 正則および非正則分布の場合の情報量

非正則分布を含むいくつかの具体的な分布において、前節で定義した情報量を通して統計量の性質を論じる。

例題 2.1.（正規分布の場合）. 各 $j=1,2$ に対して、$X_1,\cdots,X_n$ がたが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。各 X_i について、2つの分布 $N(\mu_j,\sigma_j^2)$ $(j=1,2)$ の間の情報量を前節で定義したものとすれば、

$$I_{X_i}(1,2)=4\log\left(\frac{\sigma_1^2+\sigma_2^2}{2\sigma_1\sigma_2}\right)+\frac{2(\mu_1-\mu_2)^2}{\sigma_1^2+\sigma_2^2}$$

となり、$(X_1,\cdots,X_n)$ について上の2つの分布の間の情報量は $I_{(X_1,\cdots,X_n)}(1,2)=nI_{X_i}(1,2)$ になる。そこで $T_1=\bar{X}=\sum_{i=1}^{n}X_i/n$，$T_2=\sum_{i=1}^{n}(X_i-\bar{X})^2$ とするとき、T_1, T_2 それぞれについて上の2つの分布間の情報量は

$$I_{T_1}(1,2)=4\log\left(\frac{\sigma_1^2+\sigma_2^2}{2\sigma_1\sigma_2}\right)+\frac{2n(\mu_1-\mu_2)^2}{\sigma_1^2+\sigma_2^2},$$

$$I_{T_2}(1,2)=4(n-1)\log\frac{\sigma_1^2+\sigma_2^2}{2\sigma_1\sigma_2}$$

となり、

$$I_{T_1}(1,2)+I_{T_2}(1,2)=nI_{X_i}(1,2)=I_{(X_1,\cdots,X_n)}(1,2)$$

になる。この等式は、(T_1,T_2) が (μ,σ^2) に対する十分統計量になっていることからも導かれる。

例題 2.2(t分布の場合). 各 $j=1,2$ に対して、$X_1, \cdots, X_n$ をたがいに独立にいずれも自由度3の t分布、すなわち密度関数

$$f(x-\theta_j)=\frac{C}{\{1+(x-\theta_j)^2\}^2}$$

をもつ分布に従う確率変数とする. ただし C はある定数とする. この分布は3次以上の積率が存在しないという意味で非正則な場合の1つと考えられる. 各 X_i について、2つの分布 $f(x-\theta_j)$ $(j=1,2)$ の間の情報量は

$$I_{X_i}(1,2)=-8\log\left\{1+\frac{1}{4}(\theta_1-\theta_2)^2\right\}$$

となるから、$I_{(X_1,\cdots,X_n)}(1,2)=nI_{X_i}(1,2)$ も求められる.

例題 3.2 (指数分布の場合). 各 $j=1,2$ に対して、$X_1,\cdots,X_n$ をたがいに独立にいずれも指数分布、すなわち密度関数

$$f(x,\theta_j,\xi_j)=\begin{cases}\dfrac{1}{\theta_j}\exp\left(-\dfrac{x-\xi_j}{\theta_j}\right) & (x\geq\xi_j),\\ 0 & (x<\xi_j)\end{cases}$$

をもつ分布に従う確率変数とする.

$\xi_1<\xi_2$ のとき、各 X_i について2つの分布 $f(x,\theta_j,\xi_j)$ $(j=1,2)$ の間の情報量は

$$I_{X_i}(1,2)=-8\log\left[\int_{\xi_2}^{\infty}\frac{1}{(\theta_1\theta_2)^{1/2}}\exp\left\{-\frac{1}{2}\left(\frac{x-\xi_1}{\theta_1}+\frac{x-\xi_2}{\theta_2}\right)\right\}dx\right]$$

$$= 4\log\frac{(\theta_1+\theta_2)^2}{4\theta_1\theta_2} + \frac{4(\xi_2-\xi_1)}{\theta_1}$$

となる。

そこで $T_1 = \min_{1\le i\le n} X_i$, $T_2 = \bar{X} - \min_{1\le i\le n} X_i$ とおくと、 nT_1 は指数分布に従い、 nT_2 は $(n-1)$ 個のたがいに独立な指数分布に従う確率変数の和として表わされる、すなわち $nT_2 = \sum_{i=1}^{n-1}(n-i)(X_{(i+1)} - X_{(i)})$ となる。ただし $X_{(1)} \le \cdots \le X_{(n)}$ とする。従って、T_1, T_2 について2つの分布 $f(x, \theta_j, \xi_j)$ $(j=1,2)$ の間の情報量は、それぞれ

$$I_{T_1}(1,2) = 4\log\frac{(\theta_1+\theta_2)^2}{4\theta_1\theta_2} + 4n\frac{\xi_2-\xi_1}{\theta_1} ,$$

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

となり、

$$I_{T_1}(1,2) + I_{T_2}(1,2) = nI_{X_i}(1,2) = I_{(X_1,\cdots,X_n)}(1,2)$$

が成り立つ。

<u>例題 3.4</u> (一様分布の場合). 各 $j=1,2$ に対して、$X_1, \cdots, X_n$ をたがいに独立にいずれも区間 $(\theta_j - (\xi_j/2), \theta_j + (\xi_j/2))$ 上の一様分布 U_j に従う確率変数とする。$\xi_1 \le \xi_2$ のとき、各 X_i について2つの分布 U_j $(j=1,2)$ の間の情報量は

$$I_{X_i}(1,2) = \begin{cases} \infty & \left(\theta_1 + \frac{\xi_1}{2} \le \theta_2 - \frac{\xi_2}{2} \text{ または } \theta_1 - \frac{\xi_1}{2} \ge \theta_2 + \frac{\xi_2}{2}\right), \\ -4\log\left\{\frac{1}{\xi_1\xi_2}\left(\frac{\xi_1+\xi_2}{2} + \theta_1 - \theta_2\right)^2\right\} & \left(\theta_1 - \frac{\xi_1}{2} < \theta_2 - \frac{\xi_2}{2} < \theta_1 + \frac{\xi_1}{2} < \theta_2 + \frac{\xi_2}{2}\right), \end{cases}$$

$$\begin{cases} -4\log\left\{\frac{1}{\xi_1\xi_2}\left(\frac{\xi_1+\xi_2}{2}+\theta_2-\theta_1\right)^2\right\} & \left(\theta_2-\frac{\xi_2}{2}<\theta_1-\frac{\xi_1}{2}<\theta_2+\frac{\xi_2}{2}<\theta_1+\frac{\xi_1}{2}\right), \\ -4\log\frac{\xi_1}{\xi_2} & \left(\theta_2-\frac{\xi_2}{2}<\theta_1-\frac{\xi_1}{2}<\theta_1+\frac{\xi_1}{2}<\theta_2+\frac{\xi_2}{2}\right) \end{cases}$$

となる．

次に $\xi_1=\xi_2=\xi$ のとき、$T_1=\max_{1\le i\le n}X_i-\min_{1\le i\le n}X_i$ ，$T_2=(\min_{1\le i\le n}X_i+\max_{1\le i\le n}X_i)/2$ について考える．T_1 に対して 2つの分布 U_j $(j=1,2)$ の間の情報量は $I_{T_1}(1,2)=0$ であり、

$$I_{(X_1,\cdots,X_n)}(1,2)=nI_{X_i}(1,2)=\begin{cases} -8n\log\left(1-\frac{|\theta_1-\theta_2|}{\xi}\right) & (|\theta_1-\theta_2|<\xi), \\ 0 & (|\theta_1-\theta_2|\ge\xi) \end{cases}$$

である．また T_2 に対して

$$I_{T_2}(1,2)=-8\log\left\{\int_{\theta_1-\frac{\xi}{2}}^{\theta_2+\frac{\xi}{2}}\frac{n}{\xi^n}\left(\xi-2|t_2-\theta_1|\right)^{\frac{n-1}{2}}\left(\xi-2|t_2-\theta_2|\right)^{\frac{n-1}{2}}dt_2\right\}$$

$$(0<|\theta_1-\theta_2|<\xi)$$

となるから、$\Delta/n=|\theta_2-\theta_1|$ とおいて n を十分大きくとれば

$$\begin{aligned} I_{T_2}(1,2) &= -8\log\left[\int_{-\infty}^{\infty}\frac{1}{\xi}\exp\left\{-\frac{1}{\xi}\left(|t_2|+|t_2-\Delta|\right)\right\}dt_2\right]+o(\Delta) \\ &= \frac{8\Delta}{\xi}\left\{1-\log\left(1-\frac{\Delta}{\xi}\right)\right\}+o(\Delta) \\ &= \frac{8\Delta}{\xi}+o(\Delta) \end{aligned}$$

となる．一方、十分大きな n について

$$I_{(X_1,\cdots,X_n)}(1,2)=nI_{X_i}(1,2)=\frac{8\Delta}{\xi}+o(\Delta)$$

になる．故に十分大きな n について

$$I_{(X_1,\cdots,X_n)}(1,2)=I_{T_2}(1,2)=8\Delta/\xi+o(\Delta)$$

となる．

<u>例題 2.5</u>（切断正規分布の場合）．$X_1,\cdots,X_n$ をたがいに独立にいずれも密度関数

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

をもつ分布 P_θ に従う確率変数とする．ただし c はある定数とする．$0<\theta<2$ とするとき、類似度は

$$\int_{-\infty}^{\infty}\left(\frac{dP_0}{dx}\cdot\frac{dP_\theta}{dx}\right)^{1/2}dx=e^{-\theta^2/8}P_0\{|X_1|<1-\tfrac{\theta}{2}\}$$

となり、十分小さな $\theta>0$ に対して

$$\begin{aligned}I_{X_i}(0,\theta)&=\theta^2-8\log P_0\{|X_1|<1-\tfrac{\theta}{2}\}\\ &=\theta^2-8\log\left(1-k\theta-\tfrac{1}{4}k\theta^2+o(\theta^2)\right)\end{aligned}$$

となる．ただし $k=ce^{-1/2}$ とする．

また $\bar{X}=\sum_{i=1}^{n}X_i/n$ は漸近的に正規分布 $N(\theta,(1-2k)/n)$ に従うから、十分小さな $\theta>0$ に対して

(2.1) $$I_{\bar{X}}(0,\theta)=\frac{n\theta^2}{1-2k}+\frac{1}{4}\log(1-2k)-\frac{1}{4}\log n+o(n\theta^2)$$

となる．さらに $T_1=n(X_{(1)}+1-\theta)$，$T_2=n(X_{(n)}-1-\theta)$ とおくと

(2.2) $$I_{T_1}(0,\theta/n)=4k\theta+o(\theta),\quad I_{T_2}(0,\theta/n)=4k\theta+o(\theta)$$

となる．一方、一定の n について $(X_{(1)}, X_{(n)}, \overline{X})$ は十分統計量になるから

$$I_{(X_1,\cdots,X_n)}(0,\theta)=I_{X_{(1)}}(0,\theta)+I_{X_{(n)}}(0,\theta)+I_{\overline{X}}(0,\theta)$$

が成り立つ．従って (2.1), (2.2) から、十分大きな n について上式の第 1, 2, 3 項はそれぞれ $O(n\theta)$, $O(n\theta)$, $O(n\theta^2)$ のオーダーになることがわかる．

3. 情報量による漸近相対効率

$X_1, \cdots, X_n$ をたがいに独立にいずれも密度関数 $f(x,\theta)$ をもつ分布に従う実確率変数とする．ここでは位置母数の場合、すなわち $f(x,\theta)=f(x-\theta)$ の場合について考察し、このときの累積分布関数を $F(x-\theta)$ で表わす．さらに n が奇数の場合すなわち $n=2m+1$ の場合を考えることにし、$\theta_1=\theta+\Delta$, $\theta_2=\theta-\Delta$ とおく．このとき T_0 を X_i $(i=1,\cdots,n)$ の中央値、すなわち $T_0=\mathrm{med}\,X_i$ とすれば、

$$I_{T_0}(1,2)=-4n\log\left(4F(\Delta)(1-F(\Delta))\right)+4\log Q(\Delta)+o(1)$$

となる．ただし

$$Q(\Delta)=\frac{F^2(\Delta)+\{1-F(\Delta)\}^2}{F(\Delta)(1-F(\Delta))}-\frac{f'(\Delta)}{f^2(\Delta)}\{1-2F(\Delta)\}$$

とする．また

$$I_{X_i}(1,2)=-8\log\int_{-\infty}^{\infty}f^{1/2}(x-\Delta)f^{1/2}(x+\Delta)\,dx$$

であり、この右辺を $I(\Delta)$ とおく。

一般に、標本 $(X_1, \cdots, X_n)$ に関する統計量 T の漸近相対効率を $I_T(1,2)/nI_{X_i}(1,2)$ によって定義する。

今の場合、$(X_1, \cdots, X_n)$ に関する $T_0 = \mathrm{med}\, X_i$ の漸近相対効率 (asymptotic relative efficiency 略して ARE) は

$$ARE_{T_0} = \frac{I_T(1,2)}{nI_{X_i}(1,2)} = -\frac{4}{I(\Delta)} \log\left(4F(\Delta)(1-F(\Delta))\right)$$

となり、十分小さな $\Delta > 0$ に対して

$$I(\Delta) = 4I\Delta^2 + o(\Delta^2) \qquad (I \text{ は Fisher 情報量}),$$

$$4F(\Delta)(1-F(\Delta)) = 1 - 4\left(F(\Delta) - \frac{1}{2}\right)^2 = 1 - 4f^2(0)\Delta^2 + o(\Delta^2)$$

になるから

$$ARE_{T_0} = \frac{4f^2(0)}{I} + o(1)$$

となる。これは Pitman efficiency に等しいことが分かる。

<u>例題 3.1.</u>(両側指数分布の場合). 上において密度関数 $f(x)$ を $e^{-|x|}/2$ とすれば、$F(\Delta) = 1 - \frac{1}{2}e^{-\Delta}$ $(\Delta > 0)$ となり、Fisher 情報量は $I(\Delta) = 8\{\Delta - \log(1+\Delta)\}$ になる。従って $(X_1, \cdots, X_n)$ に関する $T_0 = \mathrm{med}\, X_i$ の漸近相対効率は

$$ARE_{T_0} = \frac{\Delta - \log(2 - e^{-\Delta})}{2\{\Delta - \log(1+\Delta)\}} \to 1 \quad (\Delta \to 0)$$

となる。

References

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

Fisher, R. A. (1934). Two new properties of mathematical likelihood. Proc. Roy. Soc. London, A-144, 285-307.

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

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

Takeuchi, K. (1983). 文部省科学研究費によるシンポジウム報告.