数理解析研究所講究録
第590巻 1986年 50-51

情報圧縮と統計的検定

専修大 情報管理学科 韓 太舜

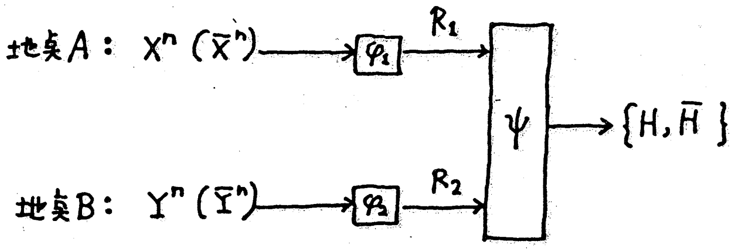
1° はじめに

統計学と情報理論は同じ情報を扱かう学問でありながら，ほとんど無関係に別個に研究され，各々の研究者の間にもそれ程の交流もなかったといってよい．前者は情報源から発生するデータを観測してパラメータの推定を行なったり検定を行なうのに対して，後者は情報源から発生する情報の圧縮限界や情報伝送の限界を調べるという，研究対象の構造の相違に主な原因があった．近年，両者を結合した研究対象と方法の重要性が指摘されているが，甘利はこれを主として"推定"の観点から研究しすでに一定の結果を提出している．本稿は，この問題を"検定"の観点から論じようとするものである．

2. 問題の枠組

$\mathcal{X}=\{1,2,\ldots,a\}$, $\mathcal{Y}=\{1,2,\ldots,b\}$ を任意の有限集合とし，$\mathcal{X}\times\mathcal{Y}$ の中に値をとる2つの確率変数 $XY\equiv(X,Y)$ と $\bar{X}\bar{Y}\equiv(\bar{X},\bar{Y})$ を考える．XY の長さ n の i.i.d.系列を $X^nY^n=(X^n,Y^n)$ と書く．$\bar{X}^n\bar{Y}^n=(X^n,Y^n)$ についても同様． 情報圧縮を考慮した仮説検定システムは図のようなものである． ここで，

仮説は H: XY であり，対立仮説は $\bar{H}$: $\bar{X}\bar{Y}$ とする．φ_1, φ_2 は符号化関



数であり，ψ は復号化関数である．地点Aで発生するデータ X^n $(\bar{X}^n)$ をすべて観測できず $\varphi_1(X^n)$ という形のデータのみを得る．地点Bで発生するデータについても $\varphi_2(Y^n)$ という形のものだけが得られる． 情報圧縮に関するレート制限は，$\log\|\varphi_i\| \leqq nR_i$ $(i=1,2)$ と表わされる． 第1種の誤り確率 $\alpha_n \leqq \varepsilon$ $(0<\varepsilon<1)$ の制限下での第2種の誤り確率の最小値を $\beta_n^*(R_1, R_2, \varepsilon)$ とする．ここに，最小値はすべての $\varphi_1, \varphi_2, \psi$ にわたる．$\theta(R_1, R_2, \varepsilon) \triangleq \varliminf_{n\to\infty}\left\{-\frac{1}{n}\log\beta_n^*(R_1, R_2, \varepsilon)\right\}$ と定義する．

3. 結果 （紙数の都合で1つの結果だけ示す）

$\mathcal{U}, \mathcal{V}$ を任意の有限集合とし，$\mathcal{U}\times\mathcal{V}$ に値をとる確率変数でマルコフ性 $U\to X\to Y\to V$ を満たすものの全体を $\mathcal{P}^*(XY)$ で表わし，$\bar{U}, \bar{V}$ を $P_{U|X}(u|x) = P_{\bar{U}|\bar{X}}(u|x)$, $P_{V|Y}(v|y) = P_{\bar{V}|\bar{Y}}(v|y)$ および $\bar{U}\to\bar{X}\to\bar{Y}\to\bar{V}$ を満たすように定める．さらに，$d(\tilde{U}\tilde{X}) = d(UX)$, $d(\tilde{V}\tilde{Y}) = d(VY)$, $d(\tilde{U}\tilde{V}) = d(UV)$ である確率変数 $\tilde{U}\tilde{X}\tilde{Y}\tilde{V}$ の全体を $\mathcal{P}^*(UVXY)$ で表わす．ただし，$d(\cdot)$ は確率分布を表わす．

定理
$$\theta(R_1, R_2, \varepsilon) \geqq \max_{\substack{UV\in\mathcal{P}^*(XY)\\ I(U:X)\leqq R_1\\ I(V:Y)\leqq R_2}} \min_{\substack{\tilde{U}\tilde{X}\tilde{Y}\tilde{V}\in\\ \mathcal{P}^*(UVXY)}} D(\tilde{U}\tilde{X}\tilde{Y}\tilde{V}\|\bar{U}\bar{X}\bar{Y}\bar{V}).$$