$$E_G[T(G;\ \boldsymbol{\chi}_n)]=E_G\left[\int g(z)\log\hat{f}(z\mid\boldsymbol{\chi}_n)dz-\frac{1}{n}\sum_{i=1}^{n}\log\hat{f}(X_i\mid\boldsymbol{\chi}_n)\right]$$

の推定とその補正が本質的となる．従って，一つの量 $T(G;\ \boldsymbol{\chi}_n)$ の変動を可能な限り制御し，精度良く期待値を推定する方法が求められる．

一般に，推定量の変動はその影響関数を通して大体の様相を把握することができる．いま，ある正則条件のもとで $T(G;\ \boldsymbol{\chi}_n)$ は，その影響関数 $IF(x;\ G)$ を用いて

$$T(G;\ \boldsymbol{\chi}_n)=\frac{1}{n}\sum_{i=1}^{n}IF(X_i;\ G)+o_p(n^{-1/2}) \tag{1}$$

と表されたとする．ブートストラップ法を適用して期待値を推定するとき，影響関数の期待値は０であることを利用すると上式は

$$\begin{aligned}T(G;\ \boldsymbol{\chi}_n)-\frac{1}{n}\sum_{i=1}^{n}IF(X_i;\ G)&=\int g(z)\log\hat{f}(z\mid\boldsymbol{\chi}_n)dz-\int g(z)\log f(z\mid\boldsymbol{\theta})dz\\&\quad+\frac{1}{n}\sum_{i=1}^{n}\log f(X_i\mid\boldsymbol{\theta})-\frac{1}{n}\sum_{i=1}^{n}\log\hat{f}(X_i\mid\boldsymbol{\chi}_n)\end{aligned} \tag{2}$$

の期待値の推定にブートストラップ法を適用することと同じことになる．(1) 式と (2) 式の違いは，(1) 式の漸近分散のオーダーが $O(1/n)$ であるのに対して，(2) 式の漸近分散のオーダーは $O(1/n^2)$ となるところにある．

情報量規準 EIC の構成に於ける (2) 式の利用は，北川 (1991) によって指摘され数値的にその良さが確かめられているが，理論的にもその有効性が示された．以上の考え方は，一般に予測誤差の推定問題において，バイアス推定の変動をおさえる一つの方法として用いることができ，ここでは特に判別分析に於ける誤判別率推定の問題に対して検討を行った．

参　考　文　献

Akaike, H. (1973). Information theory and an extention of the maximum likelihood principle, *Proceedings of 2nd International Symposium on Information Theory* (eds. B.N. Petrov and F. Csaki), 267-281, Akademiai Kiado, Budapest.

北川源四郎 (1991). 対数尤度のブートストラップについて，「時系列に関する推測の理論と応用」，共同研究レポート， No. 31, 175-179, 統計数理研究所.

第 60 回日本統計学会 (1992). 共通テーマ：ベイズ統計学の理論と応用 (II), 日本統計学会講演予稿集, 255-271.

Semimartingale Regression Model のノンパラメトリック推定

吉　田　朋　広

次の分解を持つ semimartingale regression model を考える．

$$\begin{aligned}&X_i(t)=X_i(0)+B_i(t)+M_i(t),\qquad t\in[0,\ 1],\qquad i=1,2,\dots,n,\\&B_i(t)=\int_0^t Y_i(s)b(s,\ Z_i(s))ds,\\&M_i(t)=M_i^c(t)+M_i^d(t),\\&\langle M_i^c\rangle(t)=\int_0^t Y_i(s)c^2(s,\ Z_i(s))ds,\\&M_i^d(t)=\int_0^t\int_{R^*}x[\mu_i(ds,\ dx)-\nu_i(ds,\ dx)],\\&\nu_i(ds,\ dx)=Y_i(s)n(s,\ x,\ Z_i(s))\nu_0(dx)ds.\end{aligned} \tag{1}$$

ここで，$Y_i(t)$ は indicator process で1と0の値をとる非増加 predictable process, $Z_i(t)$ は predictable な covariate process, $b(t, z)$, $c(t, z)$ は $[0, 1]\times\boldsymbol{R}$ で定義された関数，$n(s, x, z)$ は $[0, 1]\times\boldsymbol{R}^*\times\boldsymbol{R}$ で定義された関数 ($\boldsymbol{R}^*=\boldsymbol{R}-\{0\}$), M_i は2乗可積分 martingale, M_i^c は M_i の連続 martingale 部分，M_i^d は purely discontinuous な部分，μ_i は X_i の jump を表す $[0, 1]\times\boldsymbol{R}^*$ 上の random measure, ν_i は $[0, 1]\times\boldsymbol{R}^*$ 上の random measure で μ_i の compensator, ν_0 は $\boldsymbol{R}^*$ 上の既知の measure である．semimartingale regression model (1) は multiplicative intensity model や，jump のある diffusion process を含んでいる．

(X_i, Y_i, Z_i), $i=1, 2, \ldots$ は独立で同一の分布に従うとする．$t\in[0, 1]$, $z\in\boldsymbol{R}$ にたいして関数 b, c, n の汎関数

$$B(t, z)=\int_0^t b(s, z)ds,$$

$$C(t, z)=\int_0^t c^2(s, z)ds,$$

$$N(t, I, z)=\int_0^t\int_I n(s, x, z)\nu_0(dx)ds, \qquad \bar{I}\subset\boldsymbol{R}^*,$$

を時点

$$\{t_j^n;\ j=0, 1, \ldots, N_n\}, \qquad 0=t_0^n<t_1^n<\cdots<t_{N_n}^n=1,$$

で観測されたデータ

$$\{X_i(t_j^n),\ Y_i(t_j^n),\ Z_i(t_j^n);\ i=1, 2, \ldots, n,\ j=0, 1, \ldots, N_n\}$$

から推定したい．

$\boldsymbol{R}$ を幅 w_n の区間の列に分割して，z が属する区間を I_z とする．

データからつくられた random measure $\bar{\mu}_i(\cdot, z)$ を次のように定める．

$$\bar{\mu}_i([0, t]\times A, z)=\sum_{j;t_j^n\le t}1\{Z_i(t_{j-1}^n)\in I_z\}\delta_{X_i(t_j^n)-X_i(t_{j-1}^n)}(A), \qquad A\in\boldsymbol{B}(\boldsymbol{R}).$$

この random measure $\bar{\mu}_i(\cdot, z)$ から $B(t, z)$, $C(t, z)$, $N(t, I, z)$ の推定量を構成し，その一致性，漸近正規性等を示した．

統計学の基礎方程式

松縄 規

多変量観測量とそれに基づくノンパラメトリックな統計的モデルの間の統計的不確定性関係を与えた．これから，パラメトリックの場合と同様にモデル分布が満たす統計基礎方程式，それを利用した最小不確定性分布としての多変量指数分布モデルを誘導した：

ランダム行列 $\boldsymbol{A}$ から生成される k-次元列ベクトル $|(\boldsymbol{A})_k\rangle$ が測度空間 $(R^k, \boldsymbol{B}^k, \mu)$ 上で定義されているものとする．また $\langle(\boldsymbol{A})_k|=(|(\boldsymbol{A})_k\rangle)^t$ とする．$\mathscr{P}=\{P^A\}$ で多変量ノンパラメトリックモデルの分布族を表す．P^A は Λ に従属しない隠れたパラメータ Λ を含んでもよい．P^A は μ に関し絶対連続で $p(\boldsymbol{A};\Lambda)$ をその密度関数，$\boldsymbol{\Phi}(\boldsymbol{A};\Lambda)$ を観測対象の行列値関数，$\boldsymbol{\Psi}(\boldsymbol{A};\Lambda)$ でその行列値近似関数，$\boldsymbol{\Delta}\equiv\boldsymbol{\Delta}(\boldsymbol{A};\Lambda)=\boldsymbol{\Phi}(\boldsymbol{A};\Lambda)-\boldsymbol{\Psi}(\boldsymbol{A};\Lambda)$ で観測誤差を表す．モデルの持つデータの記述能力に関して $\boldsymbol{A}$ に関する性能比強度 $\boldsymbol{S}=\partial\ln p(\boldsymbol{A};\Lambda)/\partial\boldsymbol{A}$ を考える．$p(\boldsymbol{A};\Lambda)$ は観測値に基づいて構成可能であり，若干の正則条件を満たすものとする．次の；(1) 観測誤差，(2) 観測機構 (＝モデル)，(3) それらの間の相互作用；の諸量に関する平均誤差変動が存在すると仮定する：