数理解析研究所講究録
590巻 1986年 82-97

推定可能な母数のノンパラメトリック・ベイズ推定

鹿大・理 大和 元 (Hajime Yamato)

1. 序

推定可能な母数のベイズ推定を，2乗誤差に基づき，事前分布として Dirichlet invariant process (Dalal [1979]) を用いて行なう。

ノンパラメトリックなベイズ推定の為の事前分布として，Dirichlet process (Ferguson [1973]) との違いは，Dirichlet invariant process では パラメータの性質が分布に反映する点である。例えば，R^1 上でパラメータを原点対称にとれば，Dirichlet invariant process に従う分布は確率1で原点対称になる。R^2 上で直線 $y=x$ に関して対称なパラメータをとれば，Dirichlet invariant process に従う分布は確率1で R^2 上で直線 $y=x$ に関して対称になる。

Dalal [1979a, b, 1980] は Dirichlet invariant process を用いて，R^1 上の対称な分布関数および center of symmetry の推定を行っている。

ここでは，推定可能な母数の推定への応用を考える。

2. Dirichlet invariant process

$\mathscr{X}$ は d-次元 Euclidean space を表わし，$\mathscr{A}$ は d-次元 Borel class を表わすものとする。$G=\{g_1, g_2, \cdots, g_k\}$ は $\mathscr{X} \to \mathscr{X}$ への measurable transformation からなる finite group である。α は $(\mathscr{X}, \mathscr{A})$ 上の有限な measure で G の下で invariant, i.e., $\alpha(B)=\alpha(g_i B)$ for $\forall B \in \mathscr{A}$, $i=1, \ldots, k$ を満たすものとする。簡単の為に，$M=\alpha(\mathscr{X})$，$Q(\cdot)=\alpha(\cdot)/M$ で表すことにする.

$\mathscr{X}$ の mesurable partition $B_1, B_2, \cdots, B_m$ が G-不変 (invariant) であるとは，

$$B_1 \cup \cdots \cup B_m = \mathscr{X}, \quad B_i \cap B_j = \phi \ (i \neq j)$$

$$B_j = g_i B_j \quad \text{for } j=1, \cdots, m, \quad i=1, \cdots, k$$

が成り立つことである。

定義 1 (Dalal [1979 b])。 分布 P が次の (1), (2) を満たすとき，$(\mathscr{X}, \mathscr{A})$ 上でパラメータ α の Dirichet G-invariant process に従うという。

(1) P は確率 1 で，G の下で不変である。

(2) $\mathscr{X}$の任意のG－不変なmesurable partition $B_1, B_2, \ldots, B_m$に対して，$(P(B_1), \cdots, P(B_m))$ はDirichlet分布$(\alpha(B_1), \cdots, \alpha(B_m))$に従う。

分布Pが$(\mathscr{X}, \mathscr{A})$上でパラメータ$\alpha$のDirichlet G－invariant processに従うことを，簡単に $P \in DI(\alpha)$ と書くことにする。

P^*が$(\mathscr{X}, \mathscr{A})$上でパラメータ$\alpha$のDirichlet processに従うとき

$$P(A) \equiv \frac{1}{k} \sum_{j=1}^{k} P^*(g_j A) \quad \text{for} \quad A \in \mathscr{A} \tag{1}$$

によって定義されるPは $P \in DI(\alpha)$ 。

定義2（Dalal [1979 b]）． 任意m $(=1, 2, \cdots)$，及び$A_1, \cdots, A_m, C_1, \cdots, C_m$ $(\in \mathscr{A})$に対して

$$\Pr\{X_1 \in C_1, \cdots, X_m \in C_m \mid P(A_1), \cdots, P(A_m), P(C_1), \cdots, P(C_m)\}$$
$$= \prod_{i=1}^{m} P(C_i) \quad \text{a.s.}$$

が成り立つとき，$X_1, \cdots, X_n$をPからのsize nの標本という。

命題1（Dalal〔1979b〕）．　$P \in DI(\alpha)$，
XはPからのsize 1の標本とすると，任意の$g\ (\in G)$に対して，XとgXは分布Qに従う。

命題2（Dalal〔1979b〕）．　$P \in DI(\alpha)$，
$X_1, \cdots, X_n$はPからのsize nの標本とすると，$X_1, \cdots, X_n$が与えられたとき　$P \in DI(\alpha + \sum_{i=1}^{n} \delta_{X_i}^{g})$，ここで $\delta_{X_i}^{g} = \sum_{j=1}^{k} \delta_{g_j X_i} / k$　。

上の命題2により，標本$X_1, \cdots, X_n$を次の様にsequentialに得られたものと見なすことができる：

(1)　$P(\in DI(\alpha))$からsize 1の標本X_1をとる．

(2)　$P(\in DI(\alpha + \delta_{X_1}^{g}))$からsize 1の標本$X_2$をとる．

・・・

(n)　$P(\in DI(\alpha + \delta_{X_1}^{g} + \cdots + \delta_{X_{n-1}}^{g}))$からsize 1の標本$X_n$をとる。

従って，命題1から，size nの標本$X_1, \cdots, X_n$の分布を次の様に考えることができる：

(1)　$X_1 \sim Q$

(2)　$X_2 \sim \dfrac{1}{M+1}\left(MQ + \dfrac{1}{k}\sum_{j=1}^{k} \delta_{g_j X_1}\right)$, given X_1

・・・

$$(n)\quad X_n \sim \frac{1}{M+n-1}\left(MQ + \frac{1}{k}\sum_{j=1}^{k}\left(\delta_{g_j X_1} + \cdots + \delta_{g_j X_{n-1}}\right)\right)$$

$$, \text{given } X_1, \cdots, X_{n-1}。$$

3．<u>ベイズ推定</u>

$h(x)$を$(\mathfrak{X}, \mathcal{A})$上の measurable real valued fn. とし $\int_{\mathfrak{X}} h(x)dQ(x) < \infty$を満たすものとする。このとき，$P(\in DI(\alpha))$からの size 1 の標本を$X$とすると $\int_{\mathfrak{X}} h(x)dP(x) = E[h(X)|P]$ であり，命題1により

$$E\int_{\mathfrak{X}} h(x)dP(x) = Eh(X) = \int_{\mathfrak{X}} h(x)\,dQ(x) \qquad (2)$$

が成り立つ。従って，degree 1 の推定可能な母数

$$\theta_1 = \int_{\mathfrak{X}} h(x)\,dP(x)$$

に対して，$P(\in DI(\alpha))$からの size n の標本$X_1, \cdots, X_n$に基づく θ_1 のベイズ推定量は，(2)と命題2より

$$\begin{aligned}\hat{\theta}_1 &= E[\theta_1 | X_1, \cdots, X_n] \\ &= \int_{\mathfrak{X}} h(x)\, d\,\frac{\alpha(x) + \sum_{i=1}^{n}\delta_{X_i}^{g}(x)}{M+n} \\ &= \frac{M}{M+n}\int_{\mathfrak{X}} h(x)dQ(x) + \frac{1}{(M+n)k}\sum_{i=1}^{n}\sum_{j=1}^{k} h(g_j X_i).\end{aligned}$$

<u>例</u>．$(\mathfrak{X}, \mathcal{A}) = (R^1, B^1)$とし，$G = \{e, g\}$，$e(x) = x$，$g(x) = 2c - x$ for $\forall x \in R^1$とする。但し，cは定数。

任意に固定したtに対して　$h(x)=1$ $(x\leq t)$，$=0$ $(x>t)$ とおくと，　$\theta_1=F(t)$，但しFは分布Pに対応する分布関数である。$F(t)$のベイズ推定量は

$$\hat{F}(t)=\frac{M}{M+n}F_0(t)+\frac{1}{(M+n)k}\sum_{i=1}^{n}\left\{\delta_{X_i}((-\infty,t])+\delta_{2c-X_i}((-\infty,t])\right\}$$

で与えられる。ここで，$q_n=M/(M+n)$，$F_0(t)=Q((-\infty,t])$。

Dalal [1979] は$F(t)$のベイズ推定量として，積分2乗誤差を用いて，上と同じ$\hat{F}(t)$を得ている。Dalal [1979] の方法は　$\hat{F}(t)=E[F(t)\mid X_1,\cdots,X_n]$ へ，Prop. 2と(2)の特別な場合にあたる

$$EP(A)=Q(A)\quad\text{for } A\in\mathcal{A}$$

を用いている。ただし，$P\in DI(\alpha)$。

次に，degree 2の推定可能な母数のベイズ推定について考えよう（Yamato [1985]）。

補題1.　$P\in DI(\alpha)$，

$h(x,y)$: measurable real-valued ft. on $(\mathcal{X}^2,\mathcal{A}^2)$

& symmetric in x,y

$\int_{\mathcal{X}^2}h(x,y)\,dQ(x)\,dQ(y)$，$\int_{\mathcal{X}}h(x,gx)\,dQ(x)<\infty$ $(^\forall g\in G)$

このとき，

$$E\int_{\mathcal{X}^2}h(x,y)\,dP(x)\,dP(y)$$

$$=\frac{1}{M+1}\left[M\int_{\mathfrak{X}^2}h(x,y)dQ(x)dQ(y)+\frac{1}{k}\sum_{j=1}^{k}\int_{\mathfrak{X}}h(x,g_jx)dQ(x)\right].$$

（略証）　X_1, X_2 を $P(\in \mathcal{DI}(\alpha))$ からの size 2の標本とすると

$$E\int_{\mathfrak{X}^2}h(x,y)dP(x)dP(y)=EE[h(X_1,X_2)|P]$$
$$=Eh(X_1,X_2)$$
$$=EE[h(X_1,X_2)|X_1]$$

4頁の下段のように考えることにより，

$$=E\frac{1}{M+1}\left[M\int_{\mathfrak{X}}h(X_1,x_2)dQ(x_2)+\frac{1}{k}\sum_{j=1}^{k}h(X_1,g_jX_1)\right].$$

さらに，X_1 について（分布Qをもつ）期待値をとる。

定理1.　$P\in \mathcal{DI}(\alpha)$,

$X_1,\cdots,X_n$: P からの size nの標本

$$\theta_2=\int_{\mathfrak{X}^2}h(x,y)dP(x)dP(y)$$

$h(x,y)$ は補題1の条件を満たすものとする。

このとき，θ_2 のベイズ推定量$\hat{\theta}_2$は次のようになる。

$$\hat{\theta}_2 = \frac{M+n}{M+n+1}\Big[q_n^2\int_{\mathfrak{X}^2} h(x,y)\,dQ(x)\,dQ(y)$$

$$+ \frac{2q_n(1-q_n)}{nk}\sum_{i=1}^{n}\sum_{j=1}^{k}\int_{\mathfrak{X}} h(x, g_j X_i)\,dQ(x)$$

$$+ \frac{(1-q_n)^2}{n^2k^2}\sum_{i_1,i_2}\sum_{j_1,j_2} h(g_{j_1}X_{i_1}, g_{j_2}X_{i_2})\Big]$$

$$+ \frac{1}{(M+n+1)k}\Big[q_n\sum_{j=1}^{k}\int_{\mathfrak{X}} h(x, g_j x)\,dQ(x)$$

$$+ \frac{1-q_n}{nk}\sum_i\sum_{j_1,j_2} h(g_{j_1}X_i, g_{j_2}X_i)\Big] .$$

(略証)　$X_1,\cdots,X_n$ の条件の下で $P\in DI(\alpha+\sum_{i=1}^{n}\delta_{X_i}^{g})$ となることに注意して，$\hat{\theta}_2 = E[\theta_2 | X_1,\cdots,X_n]$ へ補題１を用いる。

一般には，次の様になる。

補題２．　$P\in DI(\alpha)$，

$h(x_1,\cdots,x_s)$: measurable real-valued ft. on $(\mathfrak{X}^s, \mathcal{A}^s)$ & symmetric in $x_1,\cdots,x_s$.

このとき，

$$E\int_{\mathfrak{X}^s} h(x_1,\cdots,x_s)\prod_{i=1}^{s} dP(x_i)$$

$$=\sum_C \frac{s!\,M^{\Sigma m(i)}}{k^{s-\Sigma m(i)} M^{(s)} \prod_{i=1}^{s}[m(i)!\,i^{m(i)}]}$$

$$\times\int_{\mathfrak{X}^{\Sigma m(i)}} \sum_f h(x_{11},\cdots,x_{1m(1)},x_{21},f_1x_{21},x_{22},f_2x_{22},\cdots,$$

$$x_{s1},\cdots,f_{s-\Sigma m(i)}x_{s1})\prod_{i=1}^{s}\prod_{j=1}^{m(i)}dQ(x_{ij}),$$

ただし，下式の全ての積分の存在を仮定する。また，Σ_C は $\sum_{i=1}^{s} i\,m(i)=s$ を満たす全ての s ヶの非負整数の組 $m(1),\cdots,m(s)$ についての和を表す。Σ_f はすべての $f_1,\cdots,f_{s-\Sigma m(i)}$ $(\in G)$ についての和を表す。

（略証） 補題1の方法ではなく，Dirichlet process P^* が

$$P^*(\cdot)=\sum_{j=1}^{\infty} P_j\,\delta_{V_j}(\cdot),$$

と表わされることと，3頁(1)とを用いる。ただし，$V_1,V_2,\cdots$ は $\mathfrak{X}$ の値をとる i.i.d. な r.v. で分布 Q をもつ。$P_1,P_2,\cdots$ は $M=\alpha(\mathfrak{X})$ を通してのみ α に依存する r.v. で $V_1,V_2,\cdots$ に独立，且つ $\sum_{j=1}^{\infty}P_j=1$，$P_j\geq 0\ (j=1,2,\cdots)$ を満たす (Ferguson [1973])。

積分 $\int_{\mathfrak{X}^s} h(x_1,\cdots,x_s)\prod_{i=1}^{s}dP(x_i)$ を $V_1,V_2,\cdots$ と $P_1,P_2,\cdots$ を用いて展開した上で，任意の整数 u, $r(1),\cdots,r(u)$ に対して，$t=r(1)+\cdots+r(u)$ とおくとき

$$E\sum_{j(1),\cdots,j(u)} P_{j1}^{r(1)}\cdots P_{j(u)}^{r(u)}=(r(1)-1)!\cdots(r(u)-1)!\,M^u/M^{(t)},$$

$$(M^{(t)} = M(M+1)\cdots(M+t-1))$$

が成り立つ(Yamato [1984])ことと，Qが G-invariantであることを用いる。

degree s の推定可能な母数 $\theta_s = \int_{\mathcal{X}^s} h(x_1,\cdots,x_s)\prod_{i=1}^{s} dP(x_i)$ に対して(h は $x_1,\cdots,x_s$ について対称)，

$$h_g(x_1,\cdots,x_s) = \frac{1}{k^s}\sum_{f\in G} h(f_1x_1,\cdots,f_sx_s)$$

とおくと，G-invariantなPに対して

$$\theta_s = \int_{\mathcal{X}^s} h_g(x_1,\cdots,x_s)\prod_{i=1}^{s} dP(x_i)$$

と表わされる。命題2に注意して，上式に補題2を用いると，h_g がG-invariantであることから，

定理2. $P\in DI(\alpha)$,

$X_1,\cdots,X_n$: Pからの size n の標本

h について，補題2の右辺の積分の存在を仮定する。

このとき，推定可能な母数 θ_s のベイズ推定量 $\hat{\theta}_s$ は次のようになる。

$$\hat{\theta}_s = \sum_C \frac{s!\,(M+n)^{\Sigma m(i)}}{(M+n)^{(s)}\prod_{i=1}^{s}[m(i)!\, i^{m(i)}]} \times$$

$$\int_{\mathcal{X}^{\Sigma m(i)}} h_g(x_{11},\cdots,x_{1m(1)},x_{21},x_{21},\cdots,x_{2m(2)},x_{2m(2)},\cdots,$$
$$x_{s1},\cdots,x_{s1})\Pi_{i=1}^{s}\Pi_{j=1}^{m(i)}d\hat{P}_n(x_{ij}),$$

ただし，$\hat{P}_n = q_n Q + (1-q_n)P_n^*$，$P_n^* = \frac{1}{n}\sum_{i=1}^{n}\delta_{X_i}^{g} = \frac{1}{nk}\sum_{i=1}^{n}\sum_{j=1}^{s}\delta_{g_j X_i}$．

4. 例

例1．$(\mathcal{X},\mathcal{A})=(R^1,B^1)$とし，$G=\{e,g\}$，$e(x)=x$，$g(x)=-x$ for $x\in R^1$とする。パラメータαをG－不変即ち原点に関して対称にとることにより，$P(\in DI(\alpha))$は確率1で原点に関して対称となる。degree 2の推定可能な函数

$$\Delta=\int_{R^2}|x-y|\,dP(x)\,dP(y)$$

を考えよう，これは分布Pのcoefficient of mean differenceとして知られている。$\int|x|\,dQ(x)<\infty$とすると，Δのベイズ推定量は

$$\hat{\Delta}=\frac{M+n}{M+n+1}\Big[q_n^2\int_{R^2}|x-y|\,dQ(x)\,dQ(y)$$
$$+\frac{q_n(1-q_n)}{n}\sum_{i=1}^{n}\int_{R^1}(|x-X_i|+|x+X_i|)\,dQ(x)$$

$$+\frac{(1-q_n)^2}{2n^2}\sum_{i,j}(|X_i-X_j|+|X_i+X_j|)\Big]$$

$$+\frac{1}{M+n+1}\Big[q_n\int_{R^1}|x|\,dQ(x)+\frac{1-q_n}{n}\sum_i|X_i|\Big]\ 。$$

Qを固定して，$M\to 0$としたとき，極限ベイズ推定量は

$$\Delta^{*}=\frac{1}{n(n+1)}\Big[\sum_{i<j}(|X_i-X_j|+|X_i+X_j|)+2\sum_i|X_i|\Big]$$

で与えられる。他方，Dirichlet process を用いたときの極限ベイズ推定量は

$$\Delta^{**}=\frac{2}{n(n+1)}\sum_{i<j}|X_i-X_j|$$

で与えられる（Yamato [1977]）。

例12．$(\mathcal{X},\mathcal{A})=(R^2,\mathcal{B}^2)$とし，$G=\{e,g\}$，$e(x,y)=(x,y)$ & $g(x,y)=(y,x)$ for $(x,y)\in R^2$とする。パラメータαをG－不変，即ち直線$y=x$に関して対称にとることにより，$P(\in D\mathcal{I}(\alpha))$は確率1で直線$y=x$に関して対称になる。degree 2の推定可能な母数として，分布Pの共分散

$$\theta_2=\int_{R^2}x_1x_2\,dP(x_1,x_2)-\int_R x_1\,dP(x_1,x_2)\int_R x_2\,dP(x_1,x_2)$$

のベイズ推定を考えよう。

$\int_{R^2} x_1^2 \, dQ(x_1, x_2) < \infty$ とすると，θ のベイズ推定量は

$$\hat{\theta} = \frac{M+n}{M+n+1}\Big\{ g_n^2 \,\mathrm{Cov}(Q)$$

$$+ \frac{g_n(1-g_n)}{n}\sum_{i=1}^{n}\int_{R^2}(x_1 - X_i^{(1)})(x_2 - X_i^{(2)})\, dQ(x_1, x_2)$$

$$+ \frac{(1-g_n)^2}{4n^2}\sum_{i,j}\big[(X_i^{(1)} - X_j^{(1)})(X_i^{(2)} - X_j^{(2)}) +$$

$$(X_i^{(1)} - X_j^{(2)})(X_i^{(2)} - X_j^{(1)})\big]\Big\}$$

$$- \frac{1}{2(M+n+1)}\Big[g_n \mathrm{Var}(Q) + \frac{1-g_n}{2n}\sum_i (X_i^{(1)} - X_i^{(2)})^2\Big],$$

但し，$(X_1^{(1)}, X_1^{(2)}), \cdots, (X_n^{(1)}, X_n^{(2)})$ は P からの標本を表わす。

Q を固定して，$M \to 0$ としたとき，極限ベイズ推定量は

$$\theta^* = \frac{1}{4n(n+1)}\sum_{i,j}\big[(X_i^{(1)} - X_j^{(1)})(X_i^{(2)} - X_j^{(2)}) + (X_i^{(1)} - X_j^{(2)})(X_i^{(2)} - X_j^{(1)})\big]$$

$$- \frac{1}{4n(n+1)}\sum_i (X_i^{(1)} - X_i^{(2)})^2$$

で与えられる。他方，Dirichlet process を用いたときの極限ベイズ推定量は

$$\theta^{**} = \frac{1}{2n(n+1)}\sum_{i,j}(X_i^{(1)} - X_j^{(1)})(X_i^{(2)} - X_j^{(2)}) \quad .$$

例3. 例2と同じ仮定の下で，degree 2の推定可能な
~~母数~~

$$\theta=\int_{R^4} s(x_1-y_1)s(x_2-y_2)\,dP(x_1,x_2)\,dP(y_1,y_2)$$

のベイズ推定を考えよう。ここで，$s(t)=1\ (t>0)$，$=0$ $(t=0)$，$=-1\ (t<0)$。

θのベイズ推定量は，

$$\hat{\theta}=\frac{M+n}{M+n+1}\Big\{ g_n^2\int_{R^4} s(x_1-y_1)s(x_2-y_2)\,dQ(x_1,x_2)\,dQ(y_1,y_2)$$

$$+\frac{2g_n(1-g_n)}{n}\sum_i\int_{R^2} s(x_1-X_i^{(1)})s(x_2-X_i^{(2)})\,dQ(x_1,x_2)$$

$$+\frac{(1-g_n)^2}{2n^2}\sum_{i,j}\left[s(X_i^{(1)}-X_j^{(1)})s(X_i^{(2)}-X_j^{(2)})+s(X_i^{(1)}-X_j^{(2)})s(X_i^{(2)}-X_j^{(1)})\right]\Big\}$$

$$-\frac{1}{2(M+n+1)}\Big[g_nQ\{(x_1,x_2)\mid x_1\neq x_2\}$$

$$+\frac{1-g_n}{n}\times\text{no. of}\,\{i: X_i^{(1)}\neq X_i^{(2)},\ i=1,\cdots,n\}\Big].$$

Qを固定して，$M\to 0$としたとき，極限ベイズ推定量は

$$\theta^*=\frac{1}{2n(n+1)}\sum_{i,j}\left[s(X_i^{(1)}-X_j^{(1)})s(X_i^{(2)}-X_j^{(2)})+s(X_i^{(1)}-X_j^{(2)})s(X_i^{(2)}-X_j^{(1)})\right]$$

$$-\frac{1}{2n(n+1)}\times\text{no. of}\,\{i: X_i^{(1)}\neq X_i^{(2)},\ i=1,\cdots,n\}.$$

で与えられる。他方，Dirichlet processを用いたときの極

限ベイズ推定量は

$$\theta^{**} = \frac{1}{n(n+1)} \sum_{i,j} s(X_i^{(1)} - X_j^{(1)}) s(X_i^{(2)} - X_j^{(2)})$$

で与えられる (Yamato [1977])。なお,

$$U = \frac{1}{n(n-1)} \sum_{i,j} s(X_i^{(1)} - X_j^{(1)}) s(X_i^{(2)} - X_j^{(2)})$$

は difference-sign covariance として知られている。

References

[1] Ferguson, T.S. (1973). A Bayesian analysis of some non-parametric problems. Ann. Statist. 1, 209-230.

[2] Dalal, S.R. (1979a). Nonparametric and robust Bayes estimation of location. Proc. International Symposium on Optimizing Methods in Statistics (J.S. Rustagi Ed.). New York: Academic Press.

[3] Dalal, S.R. (1979b). Dirichlet invariant process and applications to nonparametric estimation of symmetric distribution functions. Stoch. Proc. and Appl., 9, 99-107.

[4] Dalal, S.R. (1980). Non-parametric Bayes decision theory. Bayesian Statistics (J.M. Bernardo, M.H. Degroot, D.V. Lindley & A.F.M. Smith, Eds.). Valencia, Spain: University Press.

[5] Yamato, H. (1977). Relations between limiting Bayes estimates and the U-statistics for estimable parameters of degrees 2 and 3. Commun. in Statist., A, 6, 55-66.

[6] Yamato, H. (1984). Characteristic functions of means of distributiond chosen from a Dirichlet process. Ann. Probability, 12, 262-267.

[7] Yamato, H. (1985). Bayes estimation of estimable parameter with a Dirichlet invariant process. (submitted)