Bootstrap 推定量の多項式近似

―ノンパラメトリック統計学の話題から―

高　橋　一

1 はじめに

Quenouille, Tukey 等により提唱された Jackknife 法, Efron による Bootstrap 法は大型計算機の発達・普及による計算速度の飛躍的な増加と計算コストの減少によりデータ解析の強力な武器として近年その有効性と重要性を増しつつある (Efron [1979a, b]). 一方 von-Mises, Filippova, Reeds 等により研究されている統計的汎函数の "Taylor 展開" 及び漸近理論もノンパラメトリック統計学で標準的な方法となってきている (Reeds [1976], Huber [1981], Fernholz [1983]). Jackknife 法の正当性についての議論は Quenouille [1949], Tukey [1958] 以来, 数学的及び数値解析的に数多く発表されてきている, (Miller [1974], [1978]). これらの中で恐らく最もユニークなものは, 一見無関係に見えた Jackknife 法と統計的汎函数の漸近理論が実は同一の基盤に立つものであることを発見し, von-Mises の理論にもとずき, 推定量の分散及びバイアスの Jackknife 推定の漸近的一致性を証明した Jaeckel [1972] の研究であろう (Miller [1978], Huber [1981]). Bootstrap 法の有効性に関してはいくつかの具体例をモンテカルロ法により, 又有限母集団の場合は多項分布をもちい議論されている (Efron [1979a] [1982]). 又一般の母集団では, 一定の正則性の条件下で Bootstrap 法により推定された分布函数が漸近的に正規分布に収束する事等が, Bickel-Freedman [1981] や Singh [1981] らにより証明されている.

本稿では 2 節で統計的汎函数の理論にもとずく漸近理論を主に Reeds[1976] の枠組みの中で説明し，3 節では Jackknife 法及び Bootstrap 法を簡単に議論する．本論文の主要テーマは 4 節で論じる多項式近似法で，これは Jackknife 法と Bootstrap 法の中間に位置するものである．

2 統計的汎函数と漸近理論

本節では von-Mises, Filippova 等により提唱された統計的汎函数の漸近理論を簡単に説明してゆく．証明や詳しい議論は Reeds [1976], Huber [1981, ch. 3], Fernholz [1983] を参照されたい．

$X_1, \cdots\cdots, X_n$ を未知の分布 F からの大きさ $n\geq 1$ の標本，又 θ を F のパラメタとする．(後述するが θ の例としては，$E\{X_1\}$，$\mathrm{Var}\{X_1\}$ 等を考える．パラメトリック統計学とは異なり θ が F を決定するわけではない事を注意しておく．) ここで $\hat{\theta}_n=\hat{\theta}_n(X_1, \cdots\cdots, X_n)$ を θ の推定量とするとき，我々の関心は

(2. 1)
(i) $\hat{\theta}_n$ の精度，即ち $\hat{\theta}_n$ の分散の推定.
(ii) $\hat{\theta}_n$ のバイアス推定.
(iii) $\hat{\theta}_n$ の分布の推定.

等にまとめることが出来る．しかしながら，与えられた有限標本数 n のもとで上記問題に対する解答を求めることは，F が正規分布であり又 θ が平均や分散といった，非常に簡単な場合を除き現在の確率論のレベルでは多くの場合不可能である．そこで一つの近似として標本数 n を大きくしていった時の $\hat{\theta}_n$ の (分布の) 極限にもとずく議論が意味を持ってくる．以下ノンパラメトリック統計学に於ける漸近理論を展開してゆくが，まずいくつかの定義より始める．$\mathcal{M}_1$ を R^k 上で定義された分布函数のつくる十分大きな凸集合で真の分布 F 及び一点に退化した分布 $\delta_x, x\in R^k$ をすべて含むものとする，ここで

$$\delta_x\{A\}=\begin{cases} 1 & x\in A \\ 0 & x\notin A \end{cases}$$

明らかに $X_1, \cdots\cdots, X_n$ の経験分布函数 $\hat{F}(x)=n^{-1}\sum_1^n \delta_{X_i}\{(-\infty, x]\}$ は $\mathscr{M}_1$ の要素である．さらに $\mathscr{M}$ を $\mathscr{M}_1$ を含むある位相ベクトル空間とする．さて U, T を $\mathscr{M}_1$ で定義された汎函数でパラメター θ 及び推定量 $\hat{\theta}_n$ はそれぞれ $\theta=U(F), \hat{\theta}_n=T(\hat{F})$ なるものとする．ここで一般に T は n にも依存する．即ち $\hat{\theta}_n=T_n(\hat{F})$ と書くべきであるが以下推定量を定義する汎函数は n には依存しないものと仮定する，推定量 $\hat{\theta}_n$ は $\hat{F}$ を通じてのみ n に依存するのである．この様な汎函数を"統計的汎函数"と呼ぶ．

例 2.1 確率変数 X の平均 $\theta=E\{X\}$ 及び標本平均 $\bar{X}=n^{-1}\sum_{i=1}^n X_i$ は汎函数 $U=T$ を

$$T(G)=\int xdG(x)$$

で定義すれば

$$\theta=U(F)=\int xdF(x)$$

$$\bar{X}=T(\hat{F})=\int xd\hat{F}(x)=n^{-1}\sum_{i=1}^n X_1$$

となる．

例 2.2 X の母集団分散 $\theta=E\{X-EX\}^2$ 及び標本分散 $s^2=n^{-1}\sum_1{}^n(X_i-\bar{X})^2$ は汎函数 $U=T$ を

$$T(G)=\frac{1}{2}\int\int(x-y)^2dG(x)dG(y)$$

で定義すれば上と同様 $\theta=U(F)$, $s^2=T(\hat{F})$ を得る．しかしながら分散の不偏推定量 $S^2=(n-1)^{-1}\sum_{i=1}^n(X_i-\bar{X})^2$ は統計的汎函数では表わせない．

例 2.3 θ をあるパラメター（応用上多くの場合分布の位置又は尺度パラメター）とする時，θ の函数 $h(t, x)$ に関する M—推定量 $\hat{\theta}^{(M)}$ は，

$$n^{-1}\sum_{i=1}^n h(\hat{\theta}^{(M)}, X_i)=\max_{\theta}\ n^{-1}\sum_{i=1}^n h(\theta, X_i)$$

で与えられる．従って適当な正則条件下で

$$n^{-1}\sum_{i=1}^n h'(\hat{\theta}^{(M)}, X_i)=0$$

汎函数 T を

$$\int h'(T(F), x)dF(x)=0$$

で定義すれば明らかに $\hat{\theta}^{(M)}=T(\hat{F})$ が成立している.

以下簡単のため $\theta=T(F)$, $R^k=R^1$ と仮定する. 次に統計的汎函数 T の微分及び Taylor 展開を定義する, その為にまず T の定義域を $\mathcal{M}$ に拡張する. 但し任意の $c>0$ と $G\in\mathcal{M}_1$ に対し $T(cG)=T(G)$ なる様に拡張を行うものとする. さらに任意の $F, G\in\mathcal{M}_1$ と $0\leq t\leq 1$ に対し

$$F_t^{(G)}=(1-t)F+tG$$

と書くことにする. 特に $G=\hat{F}$ の時は

$$\hat{F}_t=(1-t)F+t\hat{F}$$

と書くことにしよう, 又 $L^k(\mathcal{M}, R^1)$ を $\mathcal{M}$ から R^1 への連続な k-線型変換の全体とする.

定義 2.1 (Reeds [1976] p. 151). $\mathcal{M}$ で定義された統計的汎函数 T が F で k 回連続コンパクト微分可能であるとは下記の条件 (i) (ii) (iii) を満足する $(k+1)$ 個の i-線型変換 $d^iT_F\in L^i(\mathcal{M}, R^1)$ $i=0, 1, \cdots\cdots, k$ が存在することである

(i) $d^0T_F=T(F)$

(ii) 任意の $G\in\mathcal{M}$, $i=0, 1, \cdots\cdots, k-1$ に対し

(2.2) $$\lim_{t\to 0} t^{-1}\{d^iT_{F_t^{(G)}}-d^iT_F\}=d^{i+1}T_F(G)$$

ここで収束は $\mathcal{C}$ を $\mathcal{M}$ の任意のコンパクト集合とするとき, すべての $G\in\mathcal{C}$ に対し一様.

(iii) すべての $i=0, 1, \cdots\cdots, k$ に対し d^iT_F は F の函数として連続.

次に $T(F_t^{(G)})$ の F のまわりでの汎函数的-Taylor 展開を考える. これは初等解析学における展開と同様, 定数項 $T(F)$, 線型項 $td^1T_F(G-F)$, 二次項 $\frac{1}{2}t^2d^2T_F(G-F)\cdots\cdots$ の和で $T(F+t(G-F))$ を近似する. この時初等

解析学と同様 "Taylor" 展開の剰余項

$$\mathrm{Rem}_k(F_t^{(G)})=T(F_t^{(G)})-T_k^*(F_t^{(G)})$$

(2. 3)

$$T_k^*(F_t^{(G)})=\sum_{i=0}^{k}(t^i/i!)d^iT_F(G-F)$$

を考える.

定義 2.2 (Reeds [1976], p. 56) $\mathcal{M}$ で定義された統計的汎函数 T が F で k 次のコンパクト Taylor 展開可能であるとは $(k+1)$ 個の iー線型変換 $d^iT_F\in L^i(\mathcal{M}, R^1)$ $i=0, 1, \cdots\cdots, k$ が存在して $t\to 0$ なるとき

(2. 4)

$$\mathrm{Rem}_k(F_t^{(G)})/t^k\to 0$$

がすべての $G\in\mathcal{C}$ に対し一様に成立することを云う. ここで $\mathcal{C}$ は $\mathcal{M}$ の任意のコンパクト集合で, この時 T は $\mathcal{C}_l^k$ タイプの剰余を持つという事にする.

一般に iー線型変換 d^iT_F のかたちは復雑であるが統計学的な問題に於いては, 多くの場合 d^iT_F の積分表示を仮定できる. 即ちある R^i 上の函数

$\phi^{(i)}(x_1, \cdots\cdots, x_i)=\phi^{(i)}(x_1, \cdots\cdots, x_i\,; F, T)$ が存在して

$$d^iT_F(G-F)=\int\cdots\cdots\int\phi^{(i)}(x_1, \cdots\cdots, x_i)\prod_{l=1}^{i}d(G-F)(x_l)$$

(2. 5)

$$\int\cdots\cdots\int\phi^{(i)}(x_1, \cdots\cdots, x_i)\prod_{l=1}^{i}dF(x_l)=0$$

が成立する (cf. Huber [1981] ch. 2).

付記 2.1 定義 2.1 と定義 2.2 との関係については "T が k 回連続コンパクト微分可能であれば k 次のコンパクト Taylor 展開可能である" ことが証明されている (Reeds [1976] p. 152).

付記 2.2 T が F で一階のコンパクト微分可能であるとは, (直観的には) F の任意のコンパクト近傍で T が線型汎函数で一様に近似できることをいうわけであるが, ここで実はどの様な近傍を考えるかによって種々の微分が定義される. 例えば (2. 2) の収束が任意の有界集合上で一様ならば Frechet 微分

が，又任意の G について (2.2) が成立するならば Gateaux 微分が定義される．(それぞれの場合に対応し定義 2.2 も書きなおされる．対応する剰余を $\mathcal{G}_b^k, \mathcal{G}_s^k$ と書くことにする．) 明らかに Gateaux 微分が一番弱い条件下で成立，したがって多くの統計的汎函数は Gateaux 微分可能であるが，いろいろの望ましい性質 (合成函数の微分可能性や剰余項の確率収束等) が一般に成立しない．一方 Frechet 微分は上の問題も含め多くの有用な性質を持つが，Frechet 微分可能な統計的汎函数は非常に限られている (Dieudonné [1960], ch. 8, Huber [1981], ch. 2)．Reeds によりもちいられているコンパクト微分は両者の中間に位置するもので解析学的には合成函数の微分が成立する最も弱い条件を与える．

以下 T は k 回連続コンパクト微分可能とし (主に $k=1, 2$ の場合を考える) $T(\hat{F}_t)$ の展開を考えてゆく．まず $k=1$ として

$$T(\hat{F}_t)=T(F)+t\int \psi^{(1)}(x)d\hat{F}(x)+\mathrm{Rem}_1(\hat{F}_t)$$

$$=T_1^*(\hat{F}_t)+\mathrm{Rem}_1(\hat{F}_t).$$

$t=1$ とおけば $\hat{\theta}_n=T(\hat{F})$, $\theta=T(F)$ より

$$\hat{\theta}_n=\theta+\frac{1}{n}\sum_{i=1}^{n}\psi^{(1)}(X_i)+\mathrm{Rem}_1(\hat{F}) \tag{2.6}$$

を得る．ここで $\psi^{(1)}(X_1), \cdots\cdots \psi^{(1)}(X_n)$ は真の分布 F のもとで平均 0，分散 $\sigma^2=E\{\psi^{(1)}(X_1)\}^2$ の独立な同一分布にしたがう確率変数列である．したがって $0<\sigma^2<\infty$ ならば $n\rightarrow\infty$ なるとき

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi^{(1)}(X_i)\xrightarrow{d}N(0, \sigma^2) \tag{2.7}$$

が中心極限定理より成立する．ここでもしも

$$\sqrt{n}\,\mathrm{Rem}_1(\hat{F})\xrightarrow{p}0 \qquad \text{as } n\rightarrow\infty \tag{2.8}$$

が成り立っていれば (2.6) ～ (2.8) より

(2.9) $$\sqrt{n}[\hat{\theta}_n-\theta]\xrightarrow{d}N(0,\sigma^2) \qquad \text{as } n\to\infty$$

を得る．これより $\hat{\theta}_n$ の分散の近似として

(2.10) $$\mathrm{Var}\{\theta_n\}\approx n^{-1}\sigma^2$$

をもちいる事は合理的であろう．一方 θ_n のバイアス推定に関しては T が2回連続コンパクト微分可能とする．上と同様

$$T(\hat{F}_t)=T_2{}^*(\hat{F}_t)+\mathrm{Rem}_2(\hat{F}_t)$$

$t=1$ とすれば簡単な計算により

(2.11) $$\begin{aligned}T_2{}^*(\hat{F})&=\theta+\frac{1}{n}\sum_{i=1}^{n}\phi^{(1)}(X_i)+\frac{1}{2n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}h_2(X_i,X_j)\\ h_2(x,y)&=\phi^{(2)}(x,y)-\int[\phi^{(2)}(x,t)+\phi^{(2)}(t,y)]dF(t).\end{aligned}$$

さらに真の分布 F のもとで　$E_F\{h_2(X_1,X_2)\}=0, E_F\{h_2(X_1,X_2)|X_i\}=0, i=1,$ 2. そして $E_F\{h_2(X_i,X_j)h_2(X_l,X_k)\}=E_F\{h_2{}^2(X_i,X_j)\}$ if $(i,j)=(l,h)$, $=0$ else 等の関係も計算できる．したがって $n\to\infty$ なるとき

(2.12) $$E_F\{\mathrm{Rem}_2(\hat{F})\}=O(n^{-1})$$

であれば，$n\to\infty$ なるとき

(2.13) $$E_F\{\theta_n\}=\theta+\frac{1}{2n}E_F\{h_2(X_1,X_1)\}+O(n^{-1})$$

を得る．これより θ_n のバイアス推定として

(2.14) $$\mathrm{Bias}\approx\frac{1}{2n}E_F\{h_2(X,X)\}$$

を考えることができる．もち論，推定方式としての (2.10)，(2.14) が可能となるためには (2.8) 及び (2.12) が成立しなければならない．ここが実は統計的汎函数の理論の確率論的側面を構成する部分である．

定理 2.1 (Reeds [1976] p. 80)　もしも $\mathrm{Rem}_k(\hat{F})$ が $\mathcal{G}_l{}^k$ タイプで，$\sqrt{n}(\hat{F}-F)$ がタイトである時，もしも $\mathrm{Rem}_k(\hat{F})$ が可測であれば

(2.15) $$n^{k/2}\mathrm{Rem}_k(\hat{F})\xrightarrow{p}0 \qquad \text{as } n\to\infty$$

$\sqrt{n}(\hat{F}-F)$ がタイトであるとは任意の $\varepsilon>0$ に対しあるコンパクト集合 $\mathscr{C}\subset\mathscr{M}$ が存在して $P\{\sqrt{n}(\hat{F}-F)\in\mathscr{C}\}\geq 1-\varepsilon$ となることである．さて実際に本節の方法をもちいる場合空間 $\mathscr{M}$ にいかなる位相を導入するかが問題となる．もしも非常に強い位相を導入すれば（開集合の数が多くなる．即ちコンパクト集合の数は少なくなる故）T がコンパクト微分可能になる条件はゆるまる，しかしながら $\sqrt{n}(\hat{F}-F)$ がタイトになりにくくなる．弱い位相を導入すれば逆のことが起るわけで，この trade-off 関係を満たしつつ合理的な位相を導入することは難かしい．多くの場合，ad hoc な位相が必要な様である (Reeds [1976] ch. 5, 6)．

付記 2.3 (2.6) で $t=1$ として Taylor 展開を求めているが，これは定義 2.2 と比べると一見不合理に思える．しかしながら定義 2.2 では G, F とも固定されている，一方 (2.6) では $G=\hat{F}$ で以下 $n\to\infty$ なる場合を考えている．したがって $|\hat{F}-F|=O_p(n^{-1/2})$ as $n\to\infty$ より，正式には

$$\hat{F}_t=F+\frac{t}{\sqrt{n}}\sqrt{n}(\hat{F}-F)$$
$$=F+s[\sqrt{n}(\hat{F}-F)], \qquad s=t/\sqrt{n}$$

と書くことにより $\sqrt{n}(\hat{F}-F)=O_p(1)$．又 $t=1$ は $s=1/\sqrt{n}\to 0$ と同義となり，定義 2.2 の枠組みの中で考えられる．

付記 2.4 (2.9) で一応 $\hat{\theta}_n$ の分布の近似も与えられるが，より精密な Edgeworth タイプの展開も理論的には可能である．問題は定理 2.1 の精密化で現在の所 Frechet 微分を仮定せずに行う方法は（少くとも筆者には）不明である．統計的汎函数の Edgeworth 展開については Reeds [1976] ch. 4 又は Takahashi [1984] を参照されたい．

本節を終えるにあたり具体的に $\psi^{(1)}(x)$ を計算する方法について述べておく．(2.5) より

$$A(t)=A(t;G)=T((1-t)F+tG)=T(F)+t\int\psi^{(1)}(x)dG(x)+\cdots\cdots$$

したがって

$$\frac{d}{d_t}A(t)|_{t=0}=\int\psi^{(1)}(x)dG(x) \tag{2.16}$$

ここで $G=\delta_x$ とすれば

$$\frac{d}{dt}A_x(t)|_{t=0}=\psi^{(1)}(x) \tag{2.17}$$

ただし $A_x(t)=T[(1-t)F+t\delta_x]$ $x\in R^1$

二次，三次の“導函数”についても同様に求めることができる．最後になるが $\psi^{(1)}(x)$ を Robust 統計学では influence 函数と呼ぶ (Huber [1981])．

3 Jackknife 法と Bootstrap 法

前節と同様 θ を未知の分布 F のパラメター，θ_n を標本 $X_1,\cdots\cdots,X_n$ にもとずく θ の推定量とする．本節でも問題は (2.1) (i), (ii), (iii) の考察であるが，ここでは Jackknife 法及び Bootstrap 法をもちいこれらの問題を解いてゆく．Jackknife や Bootstrap の議論にとって $\hat{\theta}_n$ が $X_1,\cdots\cdots,X_n$ の対称な函数であることは重要であるが，$\hat{\theta}_n$ が統計的汎函数で定義されるという仮定は本来必要ではない．しかしながら前節との関連上，簡単の為にある統計的汎函数により　$\hat{\theta}_n=T(\hat{F})$ と定義されると仮定する．

3.1 Jackknife 推定　$\hat{F}^{(i)}$ を標本列 $X_1,\cdots\cdots,X_{i-1},X_{i+1},\cdots\cdots X_n$ の経験分布函数，$\hat{\theta}^{(i)}=T(\hat{F}^{(i)})$ $i=1,\cdots\cdots,n$ を i 番目の仮性推定量 (pseudo estimate) とする．$\hat{\theta}^{(i)}$ は i 番目の観測値 X_i を無視し θ を推定したものである．ここで θ の Jackknife 推定量を

$$\hat{\theta}^{(J)}=\frac{1}{n}\sum_{i=1}^{n}\hat{\theta}^{(i)}$$

で，又 i 番目の (Jackknife) 仮性値 (pseudo value) を $p(i)=n\hat{\theta}_n-(n-1)\hat{\theta}_n^{(i)}$ で定義する．さて $\hat{\theta}_n$ のバイアスを $B(\hat{\theta}_n)=E_F(\hat{\theta}_n)-\theta(F)$ とす

る時 Quenouille によるバイアス推定は

$$\widehat{\mathrm{BIAS}}_J[\hat{\theta}_n]=(n-1)(\hat{\theta}^{(J)}-\hat{\theta}_n) \tag{3.1}$$

で定義される．したがって θ の Jackknife バイアス修正ずみ推定量は，

$$\begin{aligned}\tilde{\theta}&=\hat{\theta}_n-\widehat{\mathrm{BIAS}}_J[\theta_n]\\&=n\hat{\theta}_n-(n-1)\theta^{(J)}\end{aligned} \tag{3.2}$$

となる．Quenouille によれば $\tilde{\theta}$ は $\hat{\theta}_n$ より n^{-1} の項のバイアスを取りのぞく，即ち $E_n=E_F\{\hat{\theta}_n\}$ と書く時，もしも n に依存しない F の函数 $a_1(F), a_2(F)$, ……が存在して

$$E_n=\theta+\frac{1}{n}a_1(F)+\frac{1}{n^2}a_2(F)+\cdots\cdots \tag{3.3}$$

となれば，明らかに

$$E_{n-1}=\theta+\frac{1}{n-1}a_1(F)+\frac{1}{(n-1)^2}a_2(F)+\cdots\cdots$$

したがって簡単な代数計算により

$$\begin{aligned}E_F\{\tilde{\theta}\}&=nE_n-(n-1)E_{n-1}\\&=\theta-\frac{1}{n(n-1)}a_2(F)+\cdots\cdots\end{aligned}$$

を得る．より直観的な説明は Efron [1982] ch 2.3 を参照されたい．

一方 $\hat{\theta}_n$ の分散 $\mathrm{Var}(\hat{\theta}_r)-E_F\{\hat{\theta}_n-E_F\hat{\theta}_n\}^2$ の Jackknife 推定は

$$\begin{aligned}\widehat{\mathrm{VAR}}_J[\hat{\theta}]&=\frac{n-1}{n}\sum_{i=1}^{n}(\hat{\theta}^{(i)}-\hat{\theta}^{(J)})^2\\&=\frac{1}{n(n-1)}\sum_{i=1}^{n}(p_{(i)}-p_{(\cdot)})^2,\end{aligned} \tag{3.4}$$

$p_{(\cdot)}=\frac{1}{n}\sum_{i=1}^{n}p_{(i)}$ で与えられる．(Tukey [1958], Efron [1982], ch. 3, 4) Tukey による heuristic な $\widehat{\mathrm{VAR}}_J$ の正当化は $\hat{\theta}_n=\bar{X}_n$ なる時 $\widehat{\mathrm{VAR}}_J[\bar{X}_n]=[n(n-1)]^{-1}\Sigma(X_i-\bar{X})^2$ となる事実にもとずいている．この考えは下でも論ずる一般の統計的汎函数の線型近似に拡張される．さらに $\hat{\theta}_n=\bar{X}_n$ なる時の類推から Tukey は

$$(3.5)\qquad (\hat{\theta}-\theta)/\sqrt{\widehat{\mathrm{VAR}}_J[\hat{\theta}_n]} \xrightarrow{d} N(0,1)$$

を予想した.

Jackknife 法の理論的研究は数多く発表されている. Miller [1974], [1978] による二つのサーベイ論文は過去20年間に発表された主な結果及び論文を網羅している. それらの多くは上記 Tukey 予想を個々の $\hat{\theta}_n$ について証明又は反証している様に思われる. (3.5) はもち論 $p_{(i)}$ が少くとも漸近的に独立であれば成立する. この問題については以下の議論からも明らかであるが $p_{(i)}$ の定義をみるかぎり易しくはない. 一方問題を分散及びバイアスの推定に限れば比較的見通しのよい議論が Jaeckel [1972] により展開されている. 以下彼の理論を簡単に紹介しておく. 基本的には2節で展開した漸近理論にもとずく分散及びバイアスの近似式 ((2.10), (2.14)) で $\psi^{(1)}(x;F,T)$ 及び $h_2(x,y;F,T)$ をデータにもとずいて推定するのである, そのためにまず $-1\leq t\leq 1$ なる t に対し $\hat{\theta}_n^{(i)}(t)=T(\hat{F}_t^{(i)})$ を考える. ここで $\hat{F}_t^{(i)}=(1-t)\hat{F}+t\delta_{X_i}$ $i=1,\cdots\cdots,n$ (ここで $\hat{\theta}_n^{(i)}(0)=\hat{\theta}_n$, $\hat{\theta}_n^{(i)}\left(-\dfrac{1}{n-1}\right)=\hat{\theta}^{(i)}$ である). さてそれぞれの i に対し, $T(\hat{F}_t^{(i)})$ を $t=0$ で Taylor 展開すると (2次の Taylor 展開が可能と仮定して), 2節の結果より

$$\begin{aligned}
T(\hat{F}_t^{(i)}) &= T(\hat{F})+t\int\psi^{(1)}(x,\hat{F},T)d(\delta_{X_i}-\hat{F})(x)\\
(3.6)\qquad &\quad +\frac{1}{2}t^2\int\int\psi^{(2)}(x,y,\hat{F},T)d(\delta_{X_i}-\hat{F})(x)d(\delta_{X_i}-\hat{F})(y)\\
&\quad +\mathrm{Rem}_2(\hat{F}_t^{(i)}),
\end{aligned}$$

$t=-1/(n-1)$ とすれば

$$\begin{aligned}
\hat{\theta}^{(i)} &= \hat{\theta}_n-\frac{1}{n-1}\psi^{(1)}(X_i,\hat{F},T)\\
(3.7)\qquad &\quad +\frac{1}{2}\frac{1}{(n-1)^2}h_2(X_i,X_i,\hat{F},T)+\mathrm{Rem}_2(\hat{F}_t^{(i)})
\end{aligned}$$

従って (3.7) の右辺第三項以下を無視すれば

(3.8)
$$\hat{\theta}^{(i)}-\hat{\theta}_n\simeq-\frac{1}{n-1}\psi^{(1)}(X_i,\hat{F},T)$$

一方 (2.5) より $\int\psi^{(1)}(x,\hat{F},T)d\hat{F}=0$, そこで

(3.9)
$$\sum_{i=1}^{n}(\hat{\theta}^{(i)}-\hat{\theta}_n)\simeq\frac{1}{2(n-1)^2}\sum_{i=1}^{n}h_2(X_i,X_i,\hat{F},T)$$

を得る. (3.4) で $\hat{\theta}^{(J)}\cong\hat{\theta}_n$ とすれば (3.8) より

(3.10)
$$n\cdot\widehat{\mathrm{VAR}}_J[\hat{\theta}_n]\simeq\frac{1}{n-1}\sum_{i=1}^{n}[\psi^{(1)}(X_i,\hat{F},T)]^2$$

又 (3.9) より

(3.11)
$$n\cdot\widehat{\mathrm{BIAS}}_J[\hat{\theta}_n]\simeq\frac{1}{2(n-1)}\sum_{i=1}^{n}h_2(X_i,X_i,\hat{F},T)$$

即ち Jackknife の分散及びバイアス推定は, $\psi^{(1)}(x)$, $h_2(x,y)$ を $\hat{F}$ で推定することに外ならない. したがって定理 2.1 の条件下で $k=1$ の時

(3.12)
$$n\cdot\widehat{\mathrm{VAR}}_J[\theta_n]\xrightarrow{P}\frac{1}{2}E_F\{\psi^{(1)}(X_1)\}\qquad n\to\infty$$

又 $k=2$ ならば

(3.13)
$$n\cdot\widehat{\mathrm{BIAS}}_J[\hat{\theta}_n]\xrightarrow{P}\frac{1}{2}E_F\{h_2(X_1,X_1)\}\qquad n\to\infty$$

が成立する事は明らかであろう (Jaeckel [1972]).

上の議論よりもしも T が十分滑らかでないならば, Jackknife 推定量の漸近的一致性は保障されない. 一つの例がミディアンの分散推定である (Efron [1982], p. 16). さて Tukey 予想であるが $\hat{\theta}-\theta$ は基本的には

$$\frac{-1}{n-1}\sum_{i=1}^{n}\psi^{(1)}(X_i,\hat{F},T)+(\hat{\theta}^{(J)}-\theta)$$

に等しい, したがって中心極限定理が適用されるための一つの十分条件は $\psi^{(1)}(X_1,\hat{F},T)$ が, $\hat{F}$ に関し F のまわりで Taylor 展開可能で剰余項がゼロに確率収束することであろう. いずれにせよ数学的に厳密な証明は可能だが非常に面倒であろう.

3.2 Bootstrap 法　Efron [1979 a] により提唱された Bootstrap 法は概念的には Jackknife 法よりずっと簡単である．(2.1) の (i), (ii), (iii) は

$$(3.14)\quad \begin{aligned} &(\mathrm{i}) && \mathrm{Var}[\hat{\theta}_n]=E_F\{\hat{\theta}_n-E_F\hat{\theta}_n\}^2 \\ &(\mathrm{ii}) && B[\hat{\theta}_n]=E_F\{\hat{\theta}_n\}-\theta(F) \\ &(\mathrm{iii}) && \mathrm{Dist}(x)=P_F\{\hat{\theta}_n\leq x\} \end{aligned}$$

の推定ということだが，Bootstrap 推定は右辺の F を経験分布函数 $\hat{F}$ におきかえるだけのものである．

$$(3.15)\quad \begin{aligned} &(\mathrm{i}) && \widehat{\mathrm{VAR}}_{\mathrm{BOOTS}}[\hat{\theta}_n]=E_{\hat{F}}\{\hat{\theta}_n-E\hat{F}\hat{\theta}_n\}^2 \\ &(\mathrm{ii}) && \widehat{\mathrm{BIAS}}_{\mathrm{BOOTS}}[\hat{\theta}_n]=E_{\hat{F}}\{\hat{\theta}_n\}-\theta(F) \\ &(\mathrm{iii}) && \widehat{\mathrm{DIST}}_{\mathrm{BOOTS}}(x)=P_{\hat{F}}\{\hat{\theta}_n\leq x\} \end{aligned}$$

ここで F が正規分布で，又 θ が簡単なパラメター（汎函数）であれば右辺は理論的に計算できる，又 F の分布形が判らなくとも θ が比較的簡単な汎函数で定義されるのならば 2 節の方法が有効であろう．しかしながら近年多くつかわれだしたアダプティブ推定量（例えば Switzer [1972], Efron [1982] p. 28）等については適当な統計的汎函数の存在が，又存在したとしても解析可能かどうか，従って 2 節の方法はかならずしも有効とは限らない．この様に現実に数多く存在する“復雑”な状況において Efron は Bootstrap 推定量を推定する方法として Bootstrap 分布をもちいることを提唱した．Efron の方法は次の三段階より成る．

Step 1　与えられた標本より経験分布函数 $\hat{F}$ をつくる．さらに $\hat{\theta}_n=\hat{\theta}_n(X_1, \cdots\cdots, X_n)$ を計算する．

Step 2　分布 $\hat{F}$ より大きさ m の標本 $X_1^*, \cdots\cdots, X_m^*$ をとり $\hat{\theta}_m^*=\hat{\theta}_m^*(X_1^*, \cdots\cdots, X_m^*)$ を計算する．具体的には $\{X_1, \cdots\cdots, X_n\}$ の中から，くり返しをゆるして m 個の数字をランダムに選び，それらにもとずき $\hat{\theta}_m^*$ を計算する．

Step 3　Step 2 を B 回（例えば $B=100, 500, \cdots\cdots$etc.）くり返し $\hat{\theta}_m^{*(1)}, \hat{\theta}_m^{*(2)}, \cdots\cdots, \hat{\theta}_m^{*(B)}$ を得る，それらより

$$\widetilde{\mathrm{VAR}}_{\mathrm{BOOTS}}[\hat{\theta}_n]=\frac{1}{B-1}\sum_{b=1}^{B}(\hat{\theta}_m{}^{*(b)}-\hat{\theta}_m{}^{*(\cdot)})^2$$

$$\widetilde{\mathrm{BIAS}}_{\mathrm{BOOTS}}[\hat{\theta}_n]=\hat{\theta}_m{}^{*(\cdot)}-\hat{\theta}_n \tag{3.16}$$

$$\widetilde{\mathrm{DIST}}(x)_{\mathrm{BOOTS}}[\hat{\theta}_n]=\frac{1}{B}\#\{b:\hat{\theta}_m{}^{*(b)}\leq x\}$$

ここで $\hat{\theta}_m{}^{*(\cdot)}=B^{-1}\sum_1^B\hat{\theta}_m{}^{*(b)}$ (3.16) がいわゆる Bootstrap 推定量とよばれるもので混乱の恐れが無い限り以下 (3.16) を Bootstrap 推定量とよぶことにする．又応用上 $m=n$ とする事が多い故ここでもその様にしておく．Frequentist の確率の解釈によれば (3.14) の右辺の意味は，(3.16) の右辺で $\hat{F}$ を F におきかえ，$B\to\infty$ としたものに外ならない．したがって $\sup|\hat{F}(x)-F(x)|\to 0$ $(a.e.)$ as $n\to\infty$ が常に成立していることより，もしも $\hat{\theta}_n$ が十分滑らかな汎函数により定義されているのならば

$$\begin{aligned}&|\widetilde{\mathrm{VAR}}_{\mathrm{BOOTS}}[\hat{\theta}_n]-\mathrm{Var}[\hat{\theta}_n]|\to 0\\&|\widetilde{\mathrm{BIAS}}_{\mathrm{BOOTS}}[\hat{\theta}_n]-B(\hat{\theta}_n)|\to 0\qquad \text{as } n\to\infty\\&|\widetilde{\mathrm{DIST}}_{\mathrm{BOOTS}}(x)-\mathrm{Dist}(x)|\to 0\end{aligned}\tag{3.17}$$

が何らかの意味で成立していることは当然であろう．実際 Singh [1981] は $\theta=E_F(X)$ なる時，もしも F が非格子で $E|X_1|^3<\infty$ ならば $s_n{}^2=n^{-1}\sum_{i=1}^{n}(X_i-\bar{X})^2$ として $n\to\infty$ なるとき，

$$\begin{aligned}&P_F\{\sqrt{n}(\bar{X}^*-\bar{X})/s_n\leq x\}\\&\quad=\Phi(x)+\{\mu_3(1-x^2)/6\sigma^3\sqrt{n}\}\phi(x)+O(n^{-1})\end{aligned}\tag{3.18}$$

が x に関し一様に至る所で成り立つことを証明した．ここで $\sigma^2=\mathrm{Var}_F X_1$, $\mu_3=E_F(X-\theta)^3$，又 $\Phi(x),\phi(x)$ はそれぞれ標準正規分布函数，密度函数である．一方 Bickel-Freedman (1981) は，もしも T が二次の汎函数で，ある函数 $w(x,y)$ により

$$T(F)=\int\int w(x,y)dF(x)dF(y)$$

であれば，

$$\int\int w^2(x, y)dF(x)dF(y)<\infty$$

$$\int w^2(x, x)dF(x) \qquad <\infty$$

なる時

$$\sqrt{n}\{T(\hat{F}_n{}^*)-T(\hat{F})\}\xrightarrow{d}N(0, \sigma^2) \tag{3.19}$$

を証明した,

ここで $\sigma^2=4\int(\int w(x, y)dF(y))^2dF(x)-T^2(F)]$, $\hat{F}_n{}^*$ は $X_1{}^*, \cdots\cdots, X_n{}^*$ の経験分布函数.

しかしながら上の結果は実は Bootstrap 法の必要のない簡単なモデルにおける Bootstrap 法の有効性の議論で, その意味であまり面白くない. もっとも, より複雑なモデルに関しては, モンテカルロ法による数値実験に頼る外は無さそうである (Efron [1982], Diaconis-Efron [1983]).

一方 Bootstrap 法と Jackknife 法との関係については, いくつかの結果が知られている. 最も素朴には, 両者ともサンプルからのサンプリングにもとずいている. Jackknife が規則的なサンプリングを行うのに対し Bootstrap はランダム・サンプリングにもとずいている. この意味で Jackknife は Bootstrap の一次近似である. より厳密には次の定理がある.

定理 3.1 (Efron [1982] ch. 6) T をある統計的汎函数とする, 又 $\hat{F}^*$ を $\hat{F}$ からのランダム・サンプリングにもとずく経験分布函数とする.

(i) T が一階微分可能である時

$$\hat{T}_{LIN}(\hat{F}^*)=T(\hat{F})+\frac{1}{n}\sum_1^n\psi(X_i{}^*, \hat{F}, T)$$

を $T(\hat{F}^*)$ の一次近似とする. この時

$$\widehat{VAR}_J[T(\hat{F})]=\frac{n}{n-1}\widetilde{VAR}_{BOOTS}[\hat{T}_{LIN}]$$

(ii) T が二次の統計的汎函数ならば

$$\widehat{BIAS}_J[\hat{\theta}_n]=\frac{n-1}{n}\widetilde{BIAS}_{BOOTS}[\hat{\theta}_n]$$

定理の証明は Efron [1982] に詳しいが，$X_1, \cdots\cdots, X_n$ 及び $\hat{F}$ により“生成”された確率空間を考え，その中で2節の議論を展開すればいいことは容易に判るであろう．

多くの例が示す様に Bootstrap 法は Jackknife 法に比べより広い範囲で有効である（例えばミディアンの分散推定）．一方 Bootstrap で必要とされる計算量はモデル及び推定量が複雑になるにしたがい加速的に増える．又 Bootstrap 推定量の分散をいかに減らすかも問題となってきている．次節では Jackknife と Bootstrap の中間に位置する多項式近似について考えてゆく．

4　多項式近似法

前節の最後で述べた様に Bootstrap 分布を用いての推定には一般に非常に多くの計算時間が必要とされる．現在いかに安価にコンピューターを使えるといっても，多重回帰問題，Regression Tree 法 (Breiman et al [1984]) 等の問題に Bootstrap 法を適用した時等，やはり計算量を減じる方法は必要であろう．又 Bootstrap 推定の精度をいかに計るか，即ちくり返しの回数 B はどのくらいの大きさにすべきか等まだ数多くの問題が未解決のままである．一方 Jackknife 法を用いての計算量は，この様な resampling 方法の中では最少であるが，Jackknife がうまく行くためには (3.8) で見られる様に $(n-1)(\hat{\theta}_n-\hat{\theta}^{(i)})$ が T の $\hat{F}$ における influence 函数 $\psi^{(1)}(X_i, \hat{F}, T)$ のよい近似となっていなければならない．本節では Jackknife 法と Bootstrap 法の中間に位置するもの，即ち Jackknife よりも弱い条件下で有効で，しかも Bootstrap 程計算量の多くない方法について考えてゆく．基本的なアイディアは Takahashi [1982] で使われた Bernstein の多項式近似法の統計学への応用である．以下原則として前の諸節での記号をそのまま使う．本節では主に与えられた標本列 $X_1, \cdots\cdots, X_n$ とその経験的分布函数 $\hat{F}$ により構成される“Bootstrap 空

間” 上で推定量の分散推定を考える.

T を統計的汎関数とするとき

(4. 1) $$A_i(t)=T\{(1-t)\hat{F}^{(i)}+t\delta_{X_i}\} \qquad i=1,\cdots\cdots,n$$

を考える. もち論 T が $\hat{F}^{(i)}$ でコンパクト微分可能であれば $A_i(t)$ も $t=0$ で連続微分可能であるが, ここでは「$A_i(t)$ は t の函数として有界連続, ないしは有界な二次導函数を持つ」程度を仮定すれば十分である. (Bootstrap 推定をもとの確率空間の中で考える時には, より強い条件が必要であるが本節の議論は, そこまで行かない.) さて以下の分析にとって基本的なのが次に述べる Bernstein の多項式近似定理である.

補題 4.1　$\{H_{m,t}(z), m=1,2,\cdots\cdots\}$ を平均 t, 分散 $\sigma_m{}^2(t)\to 0$ as $m\to\infty$ なる分布函数列で $E_{m,t}\{f\}=\int f(z)dH_{m,t}(z)$ と書くことにする.

(a)　もしも $f(z)$ が有界連続であれば

(4. 2) $$\lim_{m\to\infty} E_{m,t}\{f\}=f(t)$$

ここで $\sigma_m{}^2(t)\to 0$ が一様となる t のすべての閉区間で (4. 2) は一様収束する.

(b)　もしも $f(z)$ が有界な二次の導函数をもてば

(4. 3) $$|E_{m,t}\{f\}-f(t)|\leq\frac{1}{2}\sigma_m{}^2(t)M_f$$

ここで $|f''(z)|\leq M_f$ 　for all z.

証明　(a) については Chebyshev の不等式の簡単な応用で証明される, 例えば Feller [1971], ch 7 を参照. (b) については, $f(z)$ を $z=t$ のまわりで展開すれば

$$f(z)=f(t)+(z-t)f'(t)+\frac{1}{2}(z-t)^2f''(t_z)$$

但だし, $|t_z-t|\leq|z-t|$

仮定より

$$\int\{f(z)-f(t)\}dH_{m,t}(z)=\frac{1}{2}\int(z-t)^2f''(z)dH_{m,t}(z)$$

$$\leq\frac{1}{2}\sigma_m{}^2(t)Mf$$

以下 $H_{m,t}(z)$ として次のものを考える．$Z_t \sim H_{mt}{}^{(t)}$ と書くとき，

$$P\left\{Z_1=\frac{k}{m}\right\}=\binom{m}{k}t^k(1-t)^{m-k}. \tag{4.4.a}$$

$k=0, 1, \cdots\cdots, m,$ なる二項分布又は

$$P\left\{Z_2=\frac{k}{m}\right\}=e^{-mk}\frac{(mt)^k}{k!} \tag{4.4.b}$$

$k=0, 1, \cdots\cdots,$ なる Poisson 分布を考える．もち論両者とも，平均は t 分散については $\mathrm{Var}\{Z_1\}=m^{-1}t(1-t)$, $\mathrm{Var}\{Z_2\}=m^{-1}t$ であり補題 4.1 の条件を満足する．

ここで $f(z)$ を $z\geq 0$ で定義された函数とすれば上記 Z_1, Z_2 のもとで $0\leq t\leq 1$ に対し

$$f(t)=\sum_{k=0}^{m}f\left(\frac{k}{m}\right)\binom{m}{k}t^k(1-t)^{m-k}+\mathrm{Rem1} \tag{4.5.a}$$

$$f(t)=e^{-mt}\sum_{k=0}^{\infty}f\left(\frac{k}{m}\right)\frac{(mt)^k}{k!}+\mathrm{Rem2} \tag{4.5.b}$$

(4.5) の右辺をそれぞれ $f(t)$ の多項式近似とよぶことにする．Rem1, Rem2 は (4.2) 又は (4.3) によりそれぞれ異なる．さて，$H_{m,t}{}^{(1)}$ で $m\to\infty, t\to 0$ しかし $mt=$ 一定　とすれば $Z_1 \xrightarrow{d} Z_2$ となる．その意味で (4.5.b) は (4.5.a) の極限である．一方上記多項式近似の意味は次の補題によっても与えられる．

補題 4.2 (Feller [1972], p. 222) 差分演算子 $\underset{h}{\Delta}$ を

$$\underset{h}{\Delta}f(t)=h^{-1}[f(t+h)-f(t)]$$

$$\underset{h}{\Delta}^k f(t)=h^{-1}[\underset{h}{\Delta}^{k-1}f(t+h)-\underset{h}{\Delta}^{k-1}f(t)], \qquad k\geq 2$$

で定義する時次の関係式が成立する

$$\sum_{k=0}^{\infty}\frac{t^k}{k!}\underset{m^{-1}}{\Delta}^k f(0)=e^{-mt}\sum_{k=0}^{\infty}f\left(\frac{k}{m}\right)\frac{(mt)^k}{k!} \tag{4.6}$$

もしも $f(z)$ が $z=0$ で k 回微分可能であれば $\underset{m^{-1}}{\Delta}^k f(0)\to(d^k/dt^k)f(t)|_{t=0}$ $m\to\infty$ が成立，すなわち (4.6) の左辺は $m\to\infty$ なる時 $f(t)$ の $t=0$ での

無限次の Taylor 展開を与える．その意味で（4.6）の右辺も $f(t)$ の“有限 Taylor 展開”を与えている．ここで重要な点は，基本的には $f(t)$ が有界連続であれば常に近似可能という点である．一方その精度はあまり良くない，補題 4.1.(b) の条件下でも，その誤差は $O(m^{-1})$ as $m\to\infty$ である．

以下（4.1）で定義した $A_i(t)$ の多項式近似を考え，それにもとずく $\hat{\theta}_n$ の分散を推定しよう．そのためにまず $A_i(t)$ は t の函数として有界連続であると仮定する．$t=n^{-1}$ での多項式近似は

$$(4.7)\qquad \hat{\theta}_n=A_i\left(\frac{1}{n}\right)=\hat{A}_i{}^{(m)}\left(\frac{1}{n}\right)+o(1)$$

as $m\to\infty$，ここで

$$\hat{A}_i{}^{(m)}(t)=\sum_{k=0}^{m}A_i\left(\frac{k}{m}\right)\binom{m}{k}t^k(1-t)^{m-k}$$

さて（4.7）の解釈であるが，これは一義的には $\hat{\theta}_n$ の X_i 方向への多項式展開で，適当な条件下で m が大きくなる時（3.7）の右辺へ収束する．又 Bootstrap の枠組の中で考えるならば $\hat{A}_i{}^{(m)}(t)$ は分布 $\hat{F}^{(i)}$ を確率 $(1-t)$ で，又 δ_{X_i} を確率 t で選ぶ様な Bootstrap 分布で θ を m 回推定した時の期待値である．したがって，$\hat{A}_i{}^{(m)}\left(\frac{1}{n}\right)$ はこの特別な Bootstrap 分布にもとずく θ の Bootstrap 推定といえる．さらに $\hat{A}_i{}^{(m)}\left(\frac{1}{n}\right)$ はおのおのの分布 $\left(1-\frac{k}{m}\right)\hat{F}^{(i)}+\frac{k}{m}\delta_{X_i}$ に対し自然な確率をあたえる．（Bootstrap 空間の中では n-カテゴリーの中から m 個とり出す多項分布が使われるが，ここで論じている分布はその周辺分布に一致している）．以上の事より $A_i\left(\frac{1}{n}\right)$の分散推定として

$$(4.8)\qquad \widetilde{VAR}_{POLY}\left[\hat{A}_i{}^{(m)}\left(\frac{1}{n}\right)\right]$$
$$=\sum_{k=0}^{m}\left[A_i\left(\frac{k}{m}\right)-A_i\left(\frac{1}{n}\right)\right]^2\binom{m}{k}\left(\frac{1}{n}\right)^k\left(1-\frac{1}{n}\right)^{m-k}$$

を得る．$\hat{F}$ が F から $O_p(1/\sqrt{n})$ のオーダーで離れている事を考えれば m をあまり大きくする事は意味をもたない．応用上は $m=n$ とする場合が多い様である．これを考慮に入れ次の多項式近似にもとずく分散推定量を定義する．

$$(4.9)\qquad \widetilde{VAR}_{POLY}[\hat{\theta}_n]=\sum_{i=1}^{n}\widetilde{VAR}_{POLY}\left[A_i{}^{(n)}\left(\frac{1}{n}\right)\right]$$

さてもしも T が線型であれば $\widetilde{VAR}_{POLY}$ は $\widehat{VAR}_J$ で $\hat{\theta}^{(J)}$ を $\hat{\theta}_n$ におきかえたものに一致する (cf. (3.8), (3.10)).

定理 4.1 T が線型統計的汎函数であれば

$$(4.10)\qquad \widetilde{VAR}_{POLY}[\hat{\theta}_n]=\frac{n-1}{n}\sum_{i=1}^{n}(\hat{\theta}^{(i)}-\hat{\theta}_n)^2$$

が成り立つ.

証明 T が線型であるから,

$$\sum_{k=0}^{n}\left[A_i\left(\frac{k}{n}\right)-\hat{\theta}_n\right]^2\binom{n}{k}\left(\frac{1}{n}\right)^k\left(1-\frac{1}{n}\right)^{n-k}$$

$$(4.11)\qquad =(\hat{\theta}^{(i)}-\hat{\theta}_n)^2+\frac{2}{n}(\hat{\theta}^{(i)}-\hat{\theta}_n)(T(\delta_{X_i})-\hat{\theta}^{(i)})$$

$$+\left(\frac{2}{n^2}-\frac{1}{n^3}\right)(T(\delta_{X_i})-\hat{\theta}^{(i)})^2$$

一方 $\hat{F}=\left(1-\frac{1}{n}\right)\hat{F}^{(i)}+\frac{1}{n}\delta_{X_i}$, より

$$\hat{\theta}^{(i)}-T(\delta_{X_i})=T(\hat{F}^{(i)}-\delta_{X_i})$$

$$=T(n\hat{F}^{(i)}-n\hat{F})=n(\hat{\theta}^{(i)}-\hat{\theta}_n)$$

これを (4.11) へ代入すれば (4.9) より定理は証明される.

一方本節の方法にもとずくバイアス推定は,

$$(4.12)\qquad \widetilde{BIAS}_{POLY}[\hat{\theta}_n]=\frac{1}{n}\sum_{i=1}^{n}\left[\hat{A}_i{}^{(m)}\left(\frac{1}{n}\right)-\hat{\theta}^{(i)}\right]$$

で与えられる. これは (3.1) で定義された Jackknife 推定が (3.3) で定義された E_n を n^{-1} の函数とみた時, その線型近似よりバイアスを推定した事を拡張したものに外ならない. ここでは E_n を n^{-1} の多項式で近似している. 定理 3.1(ii) と同様の結果は予想されるが現存未解決である.

参 考 文 献

Bickel, P. and Freedman, D. [1981]. Some asymptotic theory for the bootstrap, Ann. statist. 9, 1196—1217.

Breiman, L., Friedman, J., Olshne, R. and Stone, C. [1984] Classification and Regression Trees. Wardsworth.

Diaconis, P. and Efron, B. [1983]. Computer intensive methods in statistics, Scientific America 248, 116—130.

Dieudonné, J. [1960] Foundations of Modern Analysis. Academic Press, New York.

Efron, B. [1979 a]. Bootstrap methods: another look at the Jackknife, Ann. Statists. 7, 1—26.

Efron, B. [1979 b]. Computers and the theory of statistics: thinking the unthinkable, SIAM Rev., 21, 460—480.

Efroln, B. [1982]. The Jackknife, the Bootstrap and Other Resampling Plans, Monograph ♯38, NSF SIAM—CBMS.

Feller, W. [1971]. An Introduction to Probability Theory and Its Applications, Vol. II, 2nd Ed., John Wiley and Sons, New York.

Fernholz, L. [1983]. von-Mises Calculus for Statistical Functionals, Lecture Notes in Statistics ♯19, Academic Press.

Huber, P. [1981]. Robust Statistics, John Wiley and Sons, New York.

Jaeckel, L. [1972]. The infinitesimal Jackknife, Bell Lab Memorandum MM 72—1215—11.

Miller, R. [1974]. The Jackknife a review, Biometrika, 61, 1—17.

Miller, R. [1978]. The Jackknife: survey and applications, ARO Report 78—2.

Quenouille, M. [1949]. Approximate tests of correlation in time series, J. R. S. S. Ser. B, 11, 18—84.

Reeds, J. [1976]. On the definition of von Mises functionals, Ph. D. thesis, Harvard Univ., Cambridge, MA.

Singh, K. [1981]. On the asymptotic accuracy of Efron's Bootstrap, Ann. Statists. 9, 1187—1195.

Switzer, P. [1972]. Efficiency robustness of estimators, Proc. 6th Berkeley

Symp. 1, 283—291.

Tukey, J. [1958]. Bias and confidence in not quite large samples, abstruct, Ann. Math. Statists., 29 614.

Takahashi, H. [1985]. An application of Bernstein's theorem to some statistical problems, to appear, Fudai Keizai Ronshu, 31.

Takahashi, H. [1985]. A note on Edgeworth expansions for the von Mises functionals, to appear, J. of Multivariate Analysis.

(富山大学助教授)