

## ANOTHER RESAMPLING PLAN BASED ON THE POLYNOMIAL APPROXIMATION\*

HAJIME TAKAHASHI

### *Abstract*

We shall propose a method somewhere between the jackknife and the bootstrap. It works under weaker conditions than that of the jackknife, while it requires less numerical computation than that of the bootstrap. Our method is based on the heuristic use of the theory of functional Taylor series expansion of Reeds (1976) and the theorem of polynomial approximation to the bounded continuous function due to Bernstein.

### I. *Introduction*

Let  $X_1, \dots, X_n$  be a sequence of i.i.d. random variables on  $R^k$  ( $k \geq 1$ ) with unknown distribution function  $F$  and  $\theta = \theta(F)$  denote a parameter of interest. We shall denote  $\hat{\theta}_n = \hat{\theta}_n(X_1, \dots, X_n)$  an estimator of  $\theta$ . And we shall consider calculating the variance,  $\text{VAR}(\hat{\theta}_n) = E_F \{ \hat{\theta}_n - E_F \{ \hat{\theta}_n \} \}^2$ , and the bias,  $\text{BIAS}(\hat{\theta}_n) = E_F \{ \hat{\theta}_n \} - \theta$  of  $\hat{\theta}_n$ . Although some of the results have been announced in Takahashi (1985b), we shall restate these for the reference. We suppose that there is a statistical functional  $T$  on  $M_1$  for which  $\theta = T(F)$  and  $\hat{\theta}_n = T(\hat{F})$  for all  $n=1, 2, \dots$ , where  $\hat{F}$  denotes the empirical distribution function of  $X_1, \dots, X_n$  and  $M_1$  is a sufficiently rich convex set of distribution functions on  $R^k$  containing  $F$  and all the point mass one at  $x$ ,  $\delta_x$  ( $x \in R^k$ ). By the theory of functional Taylor series expansion, if  $T$  is compact differentiable at  $F$  (with respect to some topology on  $M_1$ ), then we may assume that there is a kernel function  $\phi^{(1)}(x, F) = \phi^{(1)}(x, F, T)$  of the derivative of  $T$  at  $F$  such that

$$(1) \quad \hat{\theta}_n = \theta + \frac{1}{n} \sum_{i=1}^n \phi^{(1)}(X_i, F) + \text{Rem } 1$$

$$(2) \quad E_F \{ \phi^{(1)}(X_1, F) \} = 0,$$

where Rem 1 is the remainder [cf. Fernholz (1983), Reeds (1976), Takahashi (1988)]. By the central limit theorem for i.i.d. random variables, if  $0 < \tau^2 = E_F \{ \phi^{(1)}(X_1, F) \}^2$  is finite and

$$(3) \quad \sqrt{n} \text{ Rem } 1 \xrightarrow{p} 0 \quad \text{as } n \rightarrow \infty.$$

Then,

---

\* The present version of the paper was supported in part by Nippon Keizai Kenkyu Josei Zaidan.

$$(4) \quad \sqrt{n} (\hat{\theta}_x - \theta) \xrightarrow{d} N(0, \tau^2) \text{ as } n \rightarrow \infty,$$

where  $N(0, \tau^2)$  denotes the normal distribution with mean 0 and variance  $\tau^2$ . The sufficient conditions for (3) are discussed by several authors [cf. Fernholz (1983), Reeds (1976)]. The above results suggests us to use  $\tau^2$  for estimating  $\text{VAR}(\hat{\theta}_n)$ . It follows that

$$(5) \quad n \cdot \widehat{\text{VAR}}_{\text{ASYM}}(\hat{\theta}_n) = \frac{1}{n} \sum_{i=1}^n [\psi^{(1)}(X_i, \hat{F})]^2$$

may be a reasonable estimator for  $\text{VAR}(\hat{\theta}_n)$ .

To obtain an estimator for  $\text{BIAS}(\hat{\theta}_n)$ , we shall suppose that  $T$  is twice compact differentiable at  $F$  with kernel function  $\psi^{(2)}(x, y, F) = \psi^{(2)}(x, y, F, T)$  for the second order derivative of  $T$  at  $F$ , for which  $\iint \psi^{(2)}(x, y, F) dF(x) dF(y) = 0$  and

$$(6) \quad \hat{\theta}_n = \theta + \frac{1}{n} \sum_{i=1}^n \psi(X_i, F) + \frac{1}{2n^2} \sum_{i=1}^n \sum_{j=1}^n h_2(X_i, X_j, F) + \text{Rem } 2,$$

where Rem 2 is the remainder and

$$h_2(x, y, F) = \psi^{(2)}(x, y, F) - \int [\psi^{(2)}(x, t, F) + \psi^{(2)}(t, y, F)] dF(t).$$

Note that  $E_F \{h_2(X_1, X_2, F)\} = 0$  holds. Thus, if

$$(7) \quad E_F \{\text{Rem } 2\} = o(n^{-1}) \text{ as } n \rightarrow \infty$$

[cf. Jaeckel (1972), Reeds (1976)], we obtain

$$(8) \quad E_F \{\hat{\theta}_n\} = \theta + \frac{1}{2n} E_F \{h_2(X_1, X_1, F)\} + o(n^{-1})$$

as  $n \rightarrow \infty$ . This suggests us to use

$$(9) \quad n \cdot \widehat{\text{BIAS}}_{\text{ASYM}}(\hat{\theta}_n) = \frac{1}{2n} \sum_{i=1}^n h_2(X_i, X_i, \hat{F})$$

as an estimator for  $\text{BIAS}(\hat{\theta}_n)$ .

The drawback of the above method is that we have to know the explicit form of  $\psi^{(1)}$  and  $\psi^{(2)}$ . But outside the text book situations, however, it is sometimes quite difficult to specify  $T$  and it is almost impossible to obtain its higher order derivatives in a closed form [see for example Switzer (1972)]. The jackknife (and the infinitesimal jackknife) and the bootstrap are alternative method for evaluating  $\text{VAR}(\hat{\theta}_n)$  and  $\text{BIAS}(\hat{\theta}_n)$ . Although the validity and the accuracy of these methods depend on the underlying distribution function  $F$  as well as the statistical functional  $T$ , they do not require the explicit knowledge of  $T$  itself. (Practically  $\hat{\theta}_n$  need not be expressed by the statistical functional for the jackknife and the bootstrap.) The jackknife estimate of  $\text{VAR}(\hat{\theta}_n)$  and  $\text{BIAS}(\hat{\theta}_n)$  are defined by

$$(10) \quad \widehat{\text{VAR}}_{\text{JACK}}(\hat{\theta}_n) = \frac{n-1}{n} \sum_{i=1}^n (\hat{\theta}_n^{(i)} - \hat{\theta}^{(J)})^2$$

$$(11) \quad \widehat{\text{BIAS}}_{\text{JACK}}(\hat{\theta}_n) = (n-1)(\hat{\theta}^{(J)} - \hat{\theta}_n),$$

where  $\hat{\theta}^{(J)} = \frac{1}{n} \sum_{i=1}^n \hat{\theta}^{(i)}$ ,  $\hat{\theta}^{(i)} = T(\hat{F}^{(i)})$  and  $\hat{F}^{(i)}$  is the empirical distribution function of  $X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n$ ,  $i=2, \dots, n-1$ ,  $\hat{F}^{(1)}$  and  $\hat{F}^{(n)}$  are the empirical distribution functions of  $X_2, \dots, X_n$  and  $X_1, \dots, X_{n-1}$  respectively. Roughly speaking the jackknife method consists of estimating  $\psi^{(1)}(x, \hat{F})$  and  $\psi^{(2)}(x, y, \hat{F})$  from the data. Indeed if  $T$  is twice compact differentiable in some neighborhood of  $F$ , then

$$(12) \quad \hat{\theta}^{(i)} = \hat{\theta}_n - \frac{1}{n-1} \psi^{(1)}(X_i, \hat{F}) + \frac{1}{2} \left( \frac{1}{n-1} \right)^2 h_2(X_i, X_i, \hat{F}) + \text{Rem } 2$$

for sufficiently large  $n$  [Takahashi (1985b)]. It follows that

$$(13) \quad \hat{\theta}^{(i)} - \hat{\theta}_n = \frac{-1}{n-1} \psi^{(1)}(X_i, \hat{F}) + o(n^{-1}) \quad (n \rightarrow \infty)$$

$$(14) \quad \sum_{i=1}^n (\hat{\theta}^{(i)} - \hat{\theta}_n) = \frac{1}{2} \left( \frac{1}{n-1} \right)^2 \sum_{i=1}^n h_2(X_i, X_i, \hat{F}) + o(n^{-2}) \quad (n \rightarrow \infty).$$

The approximation is exact in (13) and (14) if  $T$  is linear and quadratic respectively. From (10), (11), (13) and (14), it follows that

$$(15) \quad n \cdot \widehat{\text{VAR}}_{\text{JACK}}(\hat{\theta}_n) \simeq \frac{1}{n-1} \sum_{i=1}^n [\psi^{(1)}(X_i, \hat{F})]^2$$

$$(16) \quad n \cdot \widehat{\text{BIAS}}_{\text{JACK}}(\hat{\theta}_n) \simeq \frac{1}{2(n-1)} \sum_{i=1}^n h_2(X_i, X_i, \hat{F}).$$

Under some regularity conditions, the following results are well known [Efron (1982), Jaeckel (1972)].

$$(17) \quad n \cdot \widehat{\text{VAR}}_{\text{JACK}}(\hat{\theta}_n) \xrightarrow{p} E_F \{ \psi^{(1)}(X_1, F) \}^2$$

$$(18) \quad n \cdot \widehat{\text{BIAS}}_{\text{JACK}}(\hat{\theta}_n) \xrightarrow{p} \frac{1}{2} E_F \{ h_2(X_1, X_1, F) \} \quad n \rightarrow \infty.$$

The above results explain partially the reason why the jackknife estimate of variance fails when  $\hat{\theta}_n$  is the sample median. The jackknife requires the smoothness of  $T$ . On the other hand the bootstrap [Efron (1982)] works quite a large family of estimators. The price is the large amount of the numerical calculations. The comprehensive treatment of the bootstrap is found in [Efron (1982)]. (Also see Section 3.)

In this paper we shall propose a method which lies somewhere between the jackknife and the bootstrap. It works under much weaker conditions than that of the jackknife, while it requires less numerical calculations than the bootstrap. Our method is based on the polynomial approximation to the estimator  $T(\hat{F}_t^{(i)})$  at  $t = \frac{1}{n}$ , where  $\hat{F}_t^{(i)} = (1-t)\hat{F}^{(i)} + t\delta_{X_i}$ . We shall assume the continuity of  $T$  at  $F$  and utilize the polynomial approximation to the bounded continuous function by Bernstein. In Section 2 we shall state Bernstein's theorem

and its modification. Our estimator is introduced in Section 3. Some simulation results are presented in Section 4. The basic idea of this paper is found in Takahashi (1985a) which treats the finite sample space.

## II. Bernstein's Polynomial

In this section we shall present Bernstein's polynomial approximation theorem and its modification. The content of this section is not new and may be skipped except the statement of *Lemma 1*-(ii).

*Lemma 1.* Suppose  $\{H_{m,t}(z), m=1,2, \dots\}$  is a sequence of distribution functions with mean  $t$  and variance  $\sigma_m^2(t) \rightarrow 0$  as  $m \rightarrow \infty$ .

(i) If  $f(z)$  is a bounded and continuous function, then

$$(19) \quad \lim_{m \rightarrow \infty} \int f(z) dH_{m,t}(z) = f(t)$$

uniformly in the closed interval where  $\sigma_m^2(t)$  converges uniformly to 0.

(ii) If  $f(z)$  is twice differentiable with bounded second order derivative  $|f''(z)| \leq M_f$  for all  $z$ , then

$$(20) \quad \left| \int f(z) dH_{m,t}(z) - f(t) \right| \leq \frac{1}{2} \sigma_m^2(t) M_f$$

for all  $t$ .

*Proof.* The proof of the first part is a simple application of Chebyshev's inequality and is found elsewhere, see for example Feller (1971) pp. 219-220. To prove the second half, we expand  $f(z)$  into Taylor series about  $z=t$ ,

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

where  $t_z$  is the intermediate point. It follows that

$$\begin{aligned} \int [f(z) - f(t)] dH_{m,t}(z) &= \frac{1}{2} \int (z-t)^2 f''(t_z) dH_{m,t}(z) \\ &\leq \frac{1}{2} \sigma_m^2(t) M_f. \end{aligned}$$

Note that if  $f$  is linear, then the approximation is exact.

In the rest of this paper we shall use *Lemma 1* with  $H_{m,t}(z)$  a binomial distribution attaching the probability  $\binom{m}{k} t^k (1-t)^{m-k}$  to the points  $k/m$ ,  $k=0,1, \dots, m$ , whose mean and variance are  $t$  and  $t(1-t)/m$  respectively. If  $f(z)$  is a bounded continuous function on  $[0, 1]$ , then

$$(21) \quad f(t) = \sum_{k=0}^m f\left(\frac{k}{m}\right) \binom{m}{k} t^k (1-t)^{m-k} + o(1), \quad m \rightarrow \infty$$

where  $o(1)$  is of the order  $O(m^{-1})$  if  $f(z)$  is twice differentiable with bounded second order derivative for all  $t \in [0,1]$ . The meaning of the approximation (21) is found in the following lemma.

*Lemma 2.* [Feller (1971), p. 222] Let  $\Delta_h$  be a difference operator defined by

$$\begin{aligned} \Delta_h f(t) &= [f(t+h) - f(t)]/h \\ \Delta_h^k f(t) &= [\Delta_h^{k-1} f(t+h) - \Delta_h^{k-1} f(t)]/h \quad k \geq 2. \end{aligned}$$

Then

$$(22) \quad \sum_{k=0}^{\infty} \frac{t^k}{k!} \Delta_h^k f(0) = e^{-mt} \sum_{k=0}^{\infty} f\left(\frac{k}{m}\right) \frac{(mt)^k}{k!}, \quad h = m^{-1}.$$

By the Poisson approximation to the binomial distribution, if we set  $m \rightarrow \infty$  with  $mt$  being fixed, the summation on the right hand side of (21) covers to the right hand side of (22). On the other hand, if  $f(z)$  is  $k$  times differentiable at  $z=0$ , then  $\Delta_h^k f(0) \rightarrow (d^k/dz^k)f(z)|_{z=0}$  as  $h \rightarrow 0$ . Hence the left hand side of (22) is an approximation to the infinite degree Taylor series expansion of  $f(z)$  at 0. The advantage of the approximation (19) is that it holds under much weaker conditions.

### III. Polynomial Approximation

We shall keep the notation of Section 1 and write  $\hat{F}_t^{(i)} = (1-t)\hat{F}^{(i)} + t\delta_{x_i}$  for  $t \in [0,1]$ ,  $i=1, \dots, n$ . We shall also write  $A_i(t) = T(\hat{F}_t^{(i)})$ . Note that  $A_i\left(\frac{1}{n}\right) = \hat{\theta}_n$  for each  $i=1, \dots, n$ . If  $T$  is bounded continuous functional in some neighborhood of  $F$ , then  $A_i(t)$  is a bounded continuous function of  $t$  in some closed interval containing 0 and  $1/n$  for sufficiently large  $n$ ,  $[0, \Delta], \frac{1}{n} < \Delta < 1$ , say. For  $t \in [0, \Delta]$  we shall approximate  $A_i(t)$  by the Bernstein's polynomial of degree  $m$ ;

$$(23) \quad \hat{B}_i^{(m)}(t) = \sum_{k=0}^m A_i\left(\frac{k}{m}\right) \binom{m}{k} t^k (1-t)^{m-k}.$$

Our polynomial approximation to  $\hat{\theta}_n$  with respect to  $X_t$  is defined by  $\hat{B}_i^{(m)}\left(\frac{1}{n}\right)$ ;

$$(24) \quad \hat{\theta}_n = A_i\left(\frac{1}{n}\right) = \hat{B}_i^{(m)}\left(\frac{1}{n}\right) + o(1), \quad m \rightarrow \infty.$$

Here  $o(1)$  is of the order  $O(m^{-1})$  as  $m \rightarrow \infty$  if  $T$  is twice compact differentiable in the neighborhood of  $F$ . Now to estimate  $\text{VAR}(\hat{\theta}_n)$  and  $\text{BIAS}(\hat{\theta}_n)$ , Efron's bootstrap takes the sample of size  $N$  (say) from  $\hat{F}$  to calculate a bootstrap estimate  $\hat{\theta}^* = \hat{\theta}_N^*$  of  $\theta$ , and then repeat this process  $B$  (say) times to obtain  $\hat{\theta}^{*(1)}, \dots, \hat{\theta}^{*(B)}$ . Its sampling scheme is virtually the independent repetition of multinomial distribution choosing  $N$  values out of  $n$  categories

with equal probabilities. The bootstrap estimate of  $\text{VAR}(\hat{\theta}_n)$  and  $\text{BIAS}(\hat{\theta}_n)$  are defined by

$$(25) \quad \widehat{\text{VAR}}_{\text{BOOTS}}(\hat{\theta}_n) = \frac{1}{B-1} \sum_{b=1}^B (\hat{\theta}^{*(b)} - \hat{\theta}^{*(\cdot)})^2$$

$$(26) \quad \widehat{\text{BIAS}}_{\text{BOOTS}}(\hat{\theta}_n) = \hat{\theta}^{*(\cdot)} - \hat{\theta}_n,$$

where  $\hat{\theta}^{*(\cdot)} = \sum_{b=1}^B \hat{\theta}^{*(b)}/B$  [Efron (1982)]. The comparison of  $B_i^{(m)}(t)$  and  $\hat{\theta}^{*(\cdot)}$  is interesting.

It can be seen from the definition that  $B_i^{(m)}(t)$  is the mean of bootstrap estimators whose resampling scheme is based on the binomial distribution choosing  $\hat{F}^{(t)}$  and  $\delta_{x_i}$  with probability  $(1-t)$  and  $t$  respectively. In view of the bootstrap, we may define the polynomial estimate of  $\text{VAR}(\hat{\theta}_n)$  in two steps. Let us define

$$(27) \quad \widetilde{\text{VAR}}_{\text{POLY}} \left[ \hat{B}_i^{(m)} \left( \frac{1}{n} \right) \right] = \sum_{k=0}^m \left\{ A_i \left( \frac{k}{m} \right) - \hat{\theta}_n \right\}^2 \binom{m}{k} \left( \frac{1}{n} \right)^k \left( 1 - \frac{1}{n} \right)^{m-k},$$

$\widetilde{\text{VAR}}_{\text{POLY}} \left[ \hat{B}_i^{(m)} \left( \frac{1}{n} \right) \right]$  estimates the variation of  $\hat{\theta}_n$  with respect to  $X_i$   $i=1, \dots, n$ , we shall sum up all the variations to define our estimator of  $\text{VAR}(\hat{\theta}_n)$ ,

$$(28) \quad \widehat{\text{VAR}}_{\text{POLY}}^{(m)}(\hat{\theta}_n) = \sum_{i=1}^n \widetilde{\text{VAR}}_{\text{POLY}} \left[ \hat{B}_i^{(m)} \left( \frac{1}{n} \right) \right].$$

In the same way we shall define the estimator of  $\text{BIAS}(\hat{\theta}_n)$  by,

$$(29) \quad \widehat{\text{BIAS}}_{\text{POLY}}^{(m)}(\hat{\theta}_n) = \sum_{i=1}^n \left[ \hat{B}_i^{(m)} \left( \frac{1}{n} \right) - \hat{\theta}_n \right].$$

The estimators  $\widehat{\text{VAR}}_{\text{POLY}}^{(m)}(\hat{\theta}_n)$  and  $\widehat{\text{BIAS}}_{\text{POLY}}^{(m)}(\hat{\theta}_n)$  are called the polynomial estimate of variance and bias of  $\hat{\theta}_n$  respectively. They are by definition approximations for  $\widehat{\text{VAR}}_{\text{BOOTS}}(\hat{\theta}_n)$  and  $\widehat{\text{BIAS}}_{\text{BOOTS}}(\hat{\theta}_n)$ . It follows that (28) and (29) have the same limiting properties as that of the bootstrap estimators if  $T$  is continuous in some neighborhood of  $F$ . The relation to the jackknife (infinitesimal jackknife) is given in the next theorem.

*Theorem 1.* (i) If  $T$  is linear, then

$$(30) \quad \widehat{\text{VAR}}_{\text{POLY}}^{(n)}(\hat{\theta}_n) = \frac{n-1}{n} \sum_{i=1}^n (\hat{\theta}^{(i)} - \hat{\theta}_n)^2$$

(ii) If  $T$  is quadratic, then

$$(31) \quad \widehat{\text{BIAS}}_{\text{POLY}}^{(n)}(\hat{\theta}_n) = \frac{n-1}{n} \sum_{i=1}^n (\hat{\theta}^{(i)} - \hat{\theta}_n).$$

The proof of the theorem is given in the appendix.

*Remark.* From (10)~(16), it follows that  $\widehat{\text{VAR}}_{\text{POLY}}^{(n)}(\hat{\theta}_n)$  is equal to  $\widehat{\text{VAR}}_{\text{JACK}}^{(n)}(\hat{\theta}_n)$  with  $\hat{\theta}^{(j)}$  replaced by  $\hat{\theta}_n$ , and  $\widehat{\text{BIAS}}_{\text{POLY}}^{(n)}(\hat{\theta}_n) = \widehat{\text{BIAS}}_{\text{JACK}}^{(n)}(\hat{\theta}_n)$  under the appropriate conditions on  $T$  and  $F$ .

### IV. Numerical Results

In this section we shall compare the numerical results of our estimator with the others. In addition to (28) and (29) we shall also consider the following modifications of these, which may improve the numerical accuracy and stability.

$$(32) \quad \widehat{\text{VAR}}_{\text{POLY}}^{(m,t)}(\theta_n) = \sum_{i=1}^n \widehat{\text{VAR}}_{\text{POLY}}[\widehat{B}_i^{(m)}(t)]$$

$$(33) \quad \widehat{\text{BIAS}}_{\text{POLY}}^{(m,t)}(\theta_n) = \sum_{i=1}^n [B_i^{(m)}(t) - \theta_n].$$

To discuss the numerical accuracy we shall set  $\theta_n$  being the sample correlation coefficient,  $\rho_n$ . We shall consider the estimate of the standard deviation of  $\theta_n$ ,  $\sqrt{\widehat{\text{VAR}}(\theta_n)}$ .  $\sqrt{(28)}$  and  $\sqrt{(32)}$  mean the square root of (28) and (32) respectively. Tables 1 and 2 summarize the performance of several estimators applied to the law school data given in Efron (1982), p. 10.

The values of Bootstrap, Jackknife and Normal theory are taken from [Efron (1982)]. The assumption that  $\theta_n$  is expressed by the statistical functional is crucial to calculate (28), (29), (32) and (33). For example  $A_i\left(\frac{k}{m}\right)$  may be calculated from the data  $X_1, \dots, X_1, X_2, \dots, X_i, \dots, X_i, \dots, X_n$   $i=1, \dots, n$ ;  $((m-k)$  of  $X_j, j \neq i$  and  $(n-1) \times k$  of  $X_i$ ).

Now in Tables 1 and 2,  $\sqrt{(28)}$  and (29) outperform  $\sqrt{(32)}$  and (33) in the sense that the formers are more close to the Bootstrap estimates. The situation is reversed in Table 3 which gives a comparative Monte Carlo study of  $\sqrt{(28)}$ ,  $\sqrt{(32)}$ , the bootstrap, the jackknife, the infinitesimal jackknife and the normal theory estimate of  $\text{VAR}(\theta_n)$ . It is interesting to see that  $\sqrt{(28)}$  agrees with the infinitesimal jackknife quite well in both cases. On a whole,  $\sqrt{(32)}$  seems to give us a better estimate than  $\sqrt{(28)}$ . In addition to these,  $\sqrt{(32)}$  and (33) show the more numerical stability with respect to  $m$  (see Table 4).

TABLE 1

	$\sqrt{(28)}$ $m=n=15$	$\sqrt{(32)}$ $m=n=15$ $t=1/(n+n^2)$	Bootstrap $B=1,000$	Jackknife	Normal Theory
$\widehat{\text{VAR}}(\hat{\theta}_n)$	0.123	0.143	0.127	0.142	0.117

TABLE 2

	(29) $m=n=15$	(33) $m=n=15$ $t=1/(n+n^2)$	Bootstrap $B=1,000$	Jackknife	Normal Theory
$\widehat{\text{BIAS}}(\hat{\theta}_n)$	-0.005	-0.007	0.003	-0.007	-0.011

TABLE 3

A comparison of 7 methods of estimating standard deviation of  $\hat{\rho}$  and  $\hat{\phi} = \tan^{-1}\hat{\rho}$ . The Monte Carlo experiments consisted of 400 repetitions of  $X_1, \dots, X_{14} \sim$  bivariate normal with true  $\rho = 0.5$ . The true standard deviations are  $Sd(\hat{\rho}) = 0.218$ ,  $Sd(\hat{\phi}) = 0.299$ . All the components except  $\sqrt{(28)}$  and  $\sqrt{(32)}$  are taken from Efron (1982.)

	AVE	SD	AVE	SD
$\sqrt{(28)}$ $m=n=14$	0.174	0.060	0.248	0.062
$\sqrt{(32)}$ $m=n=15$ $t=1/(n+n^2)$	0.219	0.084	0.313	0.093
Bootstrap $B=512$	0.206	0.062	0.301	0.062
Jackknife	0.223	0.085	0.314	0.090
Infinitesimal Jackknife	0.175	0.058	0.244	0.052
Normal Theory	0.217	0.056	0.302	0
True	0.218		0.299	

TABLE 4

Average values of (30), (32), (31) and (33) of  $\widehat{\text{VAR}}(\hat{\rho})$ . The Monte Carlo study is conducted as in Table 3.

m	$\sqrt{(30)}$	$\sqrt{(32)}$ $t=1/(n+n^2)$	(31)	(33) $t=1/(n+n^2)$
1	0.570	0.272	-0.517	-0.050
2	0.350	0.234	0.066	-0.018
3	0.300	0.231	-0.030	-0.017
4	0.277	0.229	-0.025	-0.017
10	0.201	0.221	-0.013	-0.016
14	0.174	0.219	-0.010	-0.016

### V. Concluding Remarks

We propose the polynomial approximation mainly to justify bootstrap. Of course, as a byproduct, we have another way of performing resampling to obtain the variance as well as bias of the estimators under consideration. However, if we judge the method by the amount of calculation needed to get the estimators, the result of previous section shows that the performance of the new method is not particularly good in i.i.d. case. But the method may be useful for proving certain asymptotic results in the regression problem, for the polynomial approximation gives us a natural way to approximate influence function in this case.

It is said that regression methods may be useful when we have a missing observation. We estimate the missing values using other variables. This is very common especially when we are studying various kind of government statistics. Here the analysis is more complicated compared with the many familiar text book situations, so that jackknife method would be useful to obtain many statistical properties of the predicted values. And the polynomial



approximation would certainly give us the better result for this case and we shall study this problem using the real data in the next project.

HITOTSUBASHI UNIVERSITY

ACKNOWLEDGEMENT

The author thanks Mrs. H. Anzai for her help in programming the Monte Carlo study.

APPENDIX

We shall sketch the proof of Theorem 1. We shall suppose that  $T(\delta_{x_i})$  is always defined (this is not the case for  $\hat{\theta}_n = \hat{\rho}$ ). If not, some obvious modification should be made for each concrete case.

*Proof of (30).* By the linearity of  $T$ , it follows that

$$(A-1) \quad \sum_{k=0}^n \left[ A_i \left( \frac{k}{n} \right) - \hat{\theta}_n \right]^2 \binom{n}{k} \left( \frac{1}{n} \right)^k \left( 1 - \frac{1}{n} \right)^{n-k} \\ = (\hat{\theta}^{(i)} - \hat{\theta}_n)^2 + \frac{2}{n} (\hat{\theta}^{(i)} - \hat{\theta}_n) (T(\delta_{x_i}) - \hat{\theta}^{(i)}) + \left( \frac{2}{n^2} - \frac{1}{n^3} \right) (T(\delta_{x_i}) - \hat{\theta}^{(i)})^2.$$

It is easily seen that

$$(A-2) \quad \hat{\theta}^{(i)} - T(\delta_{x_i}) = n(\hat{\theta}^{(i)} - \hat{\theta}_n).$$

(30) follows from (A-1) and (A-2).

*Proof of (31).* Since  $T$  is twice compact differentiable at  $\hat{F}^{(i)}$ , it follows that

$$(A-3) \quad A_i(t) = T(\hat{F}^{(i)}) + t \int \phi^{(1)}(x, \hat{F}^{(i)}) d(\delta_{x_i} - \hat{F}^{(i)})(x) \\ + \frac{1}{2} t^2 \iint \phi^{(2)}(x, y, \hat{F}^{(i)}) d(\delta_{x_i} - \hat{F}^{(i)})(x) d(\delta_{x_i} - \hat{F}^{(i)})(y)$$

for each  $i=1, \dots, n$ . Since  $\hat{F}^{(i)} = \left( 1 - \frac{-1}{n-1} \right) \hat{F} + \frac{-1}{n-1} \delta_{x_i}$  and  $T$  satisfies the assumption of Lemma A-1 below, we may expand  $\phi^{(j)}(\cdot, \hat{F}^{(i)})$  at  $\hat{F}^{(i)}$  ( $j=1,2$ ). It follows that (A-3) becomes

$$\hat{\theta}^{(i)} + t \frac{n}{n-1} \phi^{(1)}(X_i, \hat{F}) - \left( t \frac{n}{(n-1)^2} - \frac{t^2}{2} \left( \frac{n}{n-1} \right)^2 \right) \hat{h}_2(X_i, X_i)$$

where  $\hat{h}_2(x, y) = h_2(x, y, \hat{F})$ .

Hence,

$$\hat{B}_i^{(m)}\left(\frac{1}{n}\right) - \hat{\theta}_n = (\hat{\theta}^{(i)} - \hat{\theta}_n) + \frac{1}{n-1} \phi^{(1)}(X_i, \hat{F}) - \frac{1}{2n(n-1)^2} \hat{h}_2(X_i, X_i).$$

It follows from (14) that

$$\sum_{i=1}^n \left[ \hat{B}_i^{(m)}\left(\frac{1}{n}\right) - \hat{\theta}_n \right] = \sum_{i=1}^n (\hat{\theta}^{(i)} - \hat{\theta}_n) - \frac{1}{2n(n-1)^2} \sum_{i=1}^n \hat{h}_2(X_i, X_i) = \left(1 - \frac{1}{n}\right) \sum_{i=1}^n (\hat{\theta}^{(i)} - \hat{\theta}_n).$$

It remains to state and prove.

*Lemma A-1.* Suppose  $T$  is three times compact differentiable at  $F$  with the kernel functions  $\phi^{(j)}$  ( $j=1, 2, 3$ ) such that

$$\begin{aligned} T(F_t) &= T(F) + t \int \phi^{(1)}(x, F) d(G-F)(x) \\ &\quad + \frac{1}{2} t^2 \iint \phi^{(2)}(x, y, F) d(G-F)(x) d(G-F)(y) \\ &\quad + \frac{1}{6} t^3 \iiint \phi^{(3)}(x, y, z, F) d(G-F)(x) d(G-F)(y) d(G-F)(z) \\ &\quad + o(t^3) \text{ as } |t| \rightarrow 0 \end{aligned}$$

uniformly in  $G$  in every compact neighborhood of  $F$ , where  $F_t = (1-t)F + tG$ . Suppose further  $\phi^{(j)}(\cdot, F)$  ( $j=1, 2$ ) is a  $(3-j)$  times compact differentiable functional at  $F$  for all  $x, y$  for which

$$\begin{aligned} \phi^{(1)}(x, F_t) &= \phi^{(1)}(x, F) + t \int \tilde{\phi}^{(2)}(x, y, F) d(G-F)(y) \\ &\quad + \frac{1}{2} t^2 \iint \tilde{\phi}^{(3)}(x, y, z, F) d(G-F)(y) d(G-F)(z) \\ &\quad + o(t^2) \end{aligned}$$

$$\phi^{(2)}(x, y, F_t) = \phi^{(2)}(x, y, F) + t \int \tilde{\phi}^{(3)}(x, y, z, F) d(G-F)(z) + o(t)$$

as  $t \rightarrow 0$  uniformly in  $G$  in every compact neighborhood of  $F$ . Then,

$$\iint \phi^{(2)}(x, y, F) d(G-F)(x) d(G-F)(y) = \iint \tilde{\phi}^{(2)}(x, y, F) d(G-F)(x) d(G-F)(y)$$

for every  $G$  in every compact neighborhood of  $F$ .

*Proof.* Let  $F_s = (1-s)F + sG$  and

$$F_t^* = (1-t)F_s + tG = (1-(s+t-st))F + (s+t-st)G.$$

We shall first expand  $T(F_t^*)$  at  $F_s$ , then expand  $T(F_s)$  and  $\phi^{(j)}(\cdot, F_s)$  at  $F$ . Compare the resulting formula with the direct expansion of  $T(F_t^*)$  at  $F$ , we have the desired result.

*REFERENCES*

- Ffron, 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* II, 2nd Ed., John Wiley & Sons, New York.
- Fernholz, L. (1983), "*von-Mises Calculus for Statistical Functionals*," Lecture Notes in Statistics 19, Academic Press.
- Jaeckel, L. (1972), "The Infinitesimal Jackknife," Bell Lab. Memorandum MM72-1215-11.
- Reeds, J. (1976), "On the Definition of von-Mises Functionals," Ph.D. thesis, Harvard University.
- Switzer, P. (1972), "Efficiency Robustness of Estimators," *Proc. 6th Berkeley Symp.* 1, pp. 283-291.
- Takahashi, H. (1985a), "An Application of Bernstein's Theorem to Some Statistical Problems," *Fudai Keizai Ronshu* 31, pp. 93-97.
- Takahashi, H. (1985b), "On the Polynomial Approximation for the Bootstrap Estimators," *Hitotsubashi Review* 94, pp. 44-65 (in Japanese).
- Takahashi, H. (1988), "A Note on Edgeworth Expansions for the von-Mises Functionals," *Journal of Multivariate Analysis* 24, pp. 56-65.