Youngnam Statistical Letters
August, 1990, Vol. 1, 59-65

Comparisons on Approximating Methods for Distribution of Sample Variance

Young-Ae Choi* and Jae-Kee Song*

ABSTRACT

The Edgeworth expansion, the Roy-Tiku method and the bootstrap method for approximating the distribution of the sample variance are compared through the Monte Carlo simulation study.

1. Introduction

A sample variance is very commonly encountered statistic, but its exact distribution is generally not known except that the underlying distribution F is a normal or a contaminated normal distribution. So it is necessary to approximate the probability distribution of the sample variance from a nonnormal population.

The purposes of this note are twofold : one is to study the approximating methods for the distribution of the sample variance such as the asymptotic method using Edgeworth expansion, the Roy-Tiku method and the bootstrap method. The other is to investigate how accurate the approximating methods are through simulation study.

Details of three approximating methods are given in Section 2. Simulations are carried out for comparing of three methods and the results are summarized in Section 3.

2. Approximating Methods

Even though the sample variance is one of the popular statistics in many fields, its exact distribution is generally not known except that the underlying distribution

* Department of Statistics, Kyungpook National University, Taegu, Korea.

F is a normal distribution or a contaminated normal distribution. So many methods for approximating the distribution of the sample variance have been proposed. Approximating methods using the Edgeworth expansion, the Roy-Tiku method and the bootstrap method are explained in this section. Let $X_1, \ldots, X_n$ be a random sample from a distribution F and $\bar{X}$ and S^2 be the sample mean and the sample variance defined by $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ and $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$, respectively. Let σ^2 be the variance of F.

(1) Edgeworth expansion

Suppose that F has finite eighth moment. Let $Y = \sqrt{n}(S^2 - \sigma^2)$. Srivastava and Chan (1989) calculate the first four cumulants ξ_1, ξ_2, ξ_3 and ξ_4 of Y in terms of the cumulants of X

$$\begin{aligned} \xi_1 &= 0 , \\ \xi_2 &= a + 2n^{-1}\kappa_2^2 + O(n^{-2}) , \\ \xi_3 &= n^{-\frac{1}{2}}b + O(n^{-\frac{3}{2}}) , \\ \xi_4 &= n^{-1}c + O(n^{-2}) . \end{aligned}$$

where κ_r stands for the r-th cumulant of X and

$$\begin{aligned} a &= \kappa_4 + 2\kappa_2^2 , \\ b &= \kappa_6 + 12\kappa_4\kappa_2 + 4\kappa_3^2 + 8\kappa_2^3 , \\ c &= \kappa_8 + 24\kappa_6\kappa_2 + 32\kappa_5\kappa_3 + 32\kappa_4^2 \\ &\quad + 144\kappa_4\kappa_2^2 + 96\kappa_3^2\kappa_2 + 48\kappa_2^4 . \end{aligned}$$

By expanding $exp\{\sum_{j=1}^{4}(it)^j \frac{\xi_j}{j!}\}$, we can approximate the characteristic function of Y. From the fact that the characteristic function uniquely determines the distribution, the distribution function of $Y/\sqrt{a}$ can be expanded for large n as

$$\begin{aligned} Pr(Y/\sqrt{a} \le z) &= \Phi(z) - n^{-\frac{1}{2}}\frac{b}{6}a^{-\frac{3}{2}}\Phi^{(3)}(z) + n^{-1}(\kappa_2^2 a^{-1}\Phi^{(2)}(z) \\ &\quad + \frac{c}{24}a^{-2}\Phi^{(4)}(z) + \frac{1}{2}(\frac{b}{6})^2 a^{-3}\Phi^{(6)}(z)) + O(n^{-\frac{3}{2}}) , \end{aligned}$$

where $\Phi(z)$ is the distribution function of the standard normal distribution and $\Phi^{(j)}(z)$ is the j-th derivative of $\Phi(z)$.

(2) The Roy-Tiku Method

Let κ_r , $r = 1,2,\ldots$, be the r-th cumulant of X and assume that $| \kappa_r/\sigma^r |$ is finite for all r. By using the Laguerre polynomials up to the fourth degree, Roy and Tiku (1962) approximated the distribution of $Q = (n-1)S^2/2\sigma^2$:

$$Pr(Q \leq v) \approx \int_0^v P_m(q) \sum_{j=0}^{k} a_j^{(m)} L_j^{(m)}(q)\, dq ,$$

where k is the number of terms in the approximation,

$$P_m(q) = \frac{1}{\Gamma(m)} q^{m-1} e^{-q} , \quad q \geq 0 , \ m \geq 0 ,$$

and $L_j^{(m)}(q)$ is the Laguerre polynomial of degree j defined by

$$L_j^{(m)}(q) = \frac{1}{j!} \sum_{i=0}^{j} \binom{j}{i} (-q)^i (\Gamma(m+j) \ / \ \Gamma(m+i)) , \quad j = 0,1,2,3,4.$$

Here $a_j^{(m)}$'s are constants defined by

$$a_j^{(m)} = \Gamma(m) \sum_{i=0}^{j} \binom{j}{i} (-1)^i E(Q^i) \ / \ \Gamma(m+i) ,$$

and the first four $a_j^{(m)}$ are given by

$$\begin{aligned}
a_0^{(m)} &= 1 , \\
a_1^{(m)} &= 0 , \\
a_2^{(m)} &= \frac{m}{(2m+1)(m+1)} \lambda_4 , \\
a_3^{(m)} &= -\frac{1}{(2m+1)(m+1)(m+2)} \left(\frac{m^2}{2m+1} \lambda_6 + (2m-1)\lambda_3^2 \right) , \\
a_4^{(m)} &= \frac{1}{(2m+1)(m+1)(m+2)(m+3)} \left(\frac{m^3}{(2m+1)^2} \lambda_8 \right. \\
&\quad \left. + \frac{8m(2m-1)}{2m+1} \lambda_5 \lambda_3 + \frac{3m^3 + 16m^2 - 2m + 1}{2m+1} \lambda_4^2 \right) ,
\end{aligned}$$

with $\lambda_r = \kappa_r/\sigma^r$.

(3) The bootstrap method

Let F_n be the empirical distribution function putting mass $\frac{1}{n}$ at each observation X_i. Let $X_1^*, \ldots, X_n^*$ be a bootstrap sample from F_n, that is, each X_i^* is independently drawn from the X_j's with probability $\frac{1}{n}$, $j = 1, 2, \ldots, n$. And let $\bar{X}^*$ and S^{2*} be the bootstrap sample mean and the bootstrap sample variance defined by $\bar{X}^* = \frac{1}{n}\sum_{i=1}^{n} X_i^*$ and $S^{2*} = \frac{1}{n-1}\sum_{i=1}^{n}(X_i^* - \bar{X}^*)^2$, respectively. Assume that there exist the fourth moment of random variable X and let $S_c^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$. Let $CDF(t) = Prob_*\{S^{2*}/S_c^2 \le t\}$ be the cumulative distribution of S^{2*}/S_c^2, where Prob$_*$ stands for the probability under the resampling. Then the bootstrap consistency is easily shown from the central limit theorem and Slutsky theorem (see Srivastava and Chan (1989)), i.e., S^{2*}/S_c^2 and S^2/σ^2 have the same asymptotic limit. We can obtain an estimate of $CDF(t)$ by $\frac{\sharp\{S^{2*}/S_c^2 \le t\}}{B}$, where B is the number of bootstrap replications. In this note, two hundred trials were done for each combination of sample size and distribution and B=200 bootstrap replications were taken for each trial.

3. A Simulation Study

To show how accurate the approximating methods described in Section 2, Monte Carlo simulations are done on a CYBER 170-835 at Kyung National University.

Let t_p be the p-th percentile of distribution function of S^2/σ^2, i.e.,

$$Pr(S^2/\sigma^2 \le t_p) = p , \quad 0 < p < 1,$$

where p is a specified value and σ^2 is the variance of F. For a given t_p, p is approximated from three methods described in Section 2, i.e., $\hat{p}_{EW}$, $\hat{p}_{RT}$, and $\hat{p}_{BT}$ are the approximated values of p by the Edgeworth expansion, the Roy-Tiku method and the bootstrap method, respectively. We take F as (i) standard normal distribution, N(0,1) (ii) uniform distribution on (0,1), U(0,1) (iii) standard exponential distribution, Exp(1) (iv) Weibull distribution with parameters 1 and 4, Weib(1,4). For each F, we generate random numbers from the appropriate IMSL subroutines. In cases (ii) $\sim$ (iv), the exact values of t_p are approximated from the 4000 replications with sample size n=100. For given t_p, p=0.05, 0.1, 0.3, 0.5, 0.7, 0.9, 0.95, and n=10, 20, 30, 50, 100, $\hat{p}_{EW}$, $\hat{p}_{RT}$, $\hat{p}_{BT}$ are compared with the specified value p and tabulated in Table 1-4.

From the results of simulation, we see that

i) The three methods perform better as sample size n gets larger.

ii) The Edgeworth expansion and the bootstrap method perform similarly in almost all distributions.

iii) The Roy-Tiku method performs very well in normal case but not good in uniform case even though sample size is large.

References

1. Efron, B. (1979). Bootstrap Methods : Another Look at the Jackknife. *Annals of Statistics* **7**, 1-26.
2. Kendall, M.G. and Stuart, A. (1969). *The Advanced Theory of Statistics, Vol.I*, : Charles W. Griffin Co., London.
3. Roy, J. and Tiku, M.L. (1962). A Laguerre Series Approximation to the Sampling Distribution of the Variance. *Sankhyā* **24**, 181-184.
4. Srivastava, M.S. and Chan, Y.M. (1989). A Comparison of Bootstrap Method and Edgeworth Expansion in Approximating the Distribution of Sample Variance - one sample and two sample cases. *Commun.Statist - Simular.*, **18(1)**, 339-361.
5. Tan, W.Y. and Wong, S.P. (1977). On the Roy-Tiku Approximation of Sample Variance From Nonnormal Universes. *J. Amer. Statist. Assoc.* **72**, 875-880.

Table 1. Comparisons of $\hat{p}_{EW}$, $\hat{p}_{RT}$ and $\hat{p}_{BT}$ when F is N(0,1)

n	p	0.05	0.10	0.30	0.50	0.70	0.90	0.95
10	$\hat{p}_{EW}$	.066	.121	.323	.506	.675	.891	.956
	$\hat{p}_{RT}$	.050	.100	.300	.500	.700	.900	.950
	$\hat{p}_{BT}$	.060	.100	.235	.432	.685	.933	.980
20	$\hat{p}_{EW}$	.055	.109	.312	.502	.688	.895	.952
	$\hat{p}_{RT}$	.050	.100	.300	.500	.700	.900	.950
	$\hat{p}_{BT}$	.052	.089	.267	.468	.705	.919	.966
30	$\hat{p}_{EW}$	.053	.106	.308	.501	.692	.886	.951
	$\hat{p}_{RT}$	.050	.100	.300	.500	.700	.900	.950
	$\hat{p}_{BT}$	.048	.085	.280	.484	.699	.912	.961
50	$\hat{p}_{EW}$	.052	.104	.305	.501	.695	.897	.950
	$\hat{p}_{RT}$	.050	.100	.300	.500	.700	.900	.950
	$\hat{p}_{BT}$	.045	.088	.286	.487	.696	.916	.956
100	$\hat{p}_{EW}$	.051	.102	.303	.500	.697	.900	.950
	$\hat{p}_{RT}$	.050	.100	.300	.500	.700	.900	.950
	$\hat{p}_{BT}$	.050	.096	.286	.502	.703	.906	.956

Table 2. Comparisons of $\hat{p}_{EW}$, $\hat{p}_{RT}$ and $\hat{p}_{BT}$ when F is U(0,1)

n	p	0.05	0.10	0.30	0.50	0.70	0.90	0.95
10	$\hat{p}_{EW}$	.337	.375	.455	.513	.570	.652	.688
	$\hat{p}_{RT}$	.105	.169	.310	.413	.515	.664	.729
	$\hat{p}_{BT}$	.334	.366	.444	.499	.556	.642	.681
20	$\hat{p}_{EW}$	.252	.303	.421	.509	.595	.717	.766
	$\hat{p}_{RT}$	-.011	.074	.279	.436	.592	.812	.901
	$\hat{p}_{BT}$	.255	.303	.419	.508	.594	.716	.767
30	$\hat{p}_{EW}$	.198	.255	.397	.507	.615	.761	.816
	$\hat{p}_{RT}$	-.088	.004	.249	.446	.642	.908	1.007
	$\hat{p}_{BT}$	.203	.260	.397	.507	.614	.758	.813
50	$\hat{p}_{EW}$	.130	.191	.362	.505	.646	.822	.880
	$\hat{p}_{RT}$	-.179	-.091	.198	.456	.714	1.033	1.133
	$\hat{p}_{BT}$	.132	.192	.364	.507	.648	.821	.880
100	$\hat{p}_{EW}$	.052	.103	.302	.503	.701	.906	.953
	$\hat{p}_{RT}$	-.249	-.209	.105	.466	.830	1.187	1.245
	$\hat{p}_{BT}$	.056	.107	.306	.506	.703	.903	.950

· $\hat{p}_{EW}$: The approximated value of the Edgeworth expansion
· $\hat{p}_{RT}$: The approximated value of the Roy–Tiku method
· $\hat{p}_{BT}$: The approximated value of the bootstrap method

Table 3. Comparisons of $\hat{p}_{EW}$, $\hat{p}_{RT}$ and $\hat{p}_{BT}$ when F is Exp(1)

n	p	0.05	0.10	0.30	0.50	0.70	0.90	0.95
10	$\hat{p}_{EW}$	.396	.454	.585	.676	.770	.892	.945
	$\hat{p}_{BT}$	.215	.245	.337	.432	.612	.814	.873
20	$\hat{p}_{EW}$	.278	.344	.501	.613	.725	.859	.910
	$\hat{p}_{BT}$	.166	.206	.332	.464	.657	.842	.905
30	$\hat{p}_{EW}$	.209	.278	.453	.582	.711	.858	.907
	$\hat{p}_{BT}$	.142	.184	.316	.468	.664	.861	.920
50	$\hat{p}_{EW}$	.126	.195	.391	.548	.705	.871	.918
	$\hat{p}_{BT}$	.088	.128	.284	.464	.687	.904	.955
100	$\hat{p}_{EW}$	.040	.092	.301	.507	.718	.909	.949
	$\hat{p}_{BT}$	.047	.080	.247	.458	.721	.934	.973

Table 4. Comparisons of $\hat{p}_{EW}$, $\hat{p}_{RT}$ and $\hat{p}_{BT}$ when F is Weib(1,4)

n	p	0.05	0.10	0.30	0.50	0.70	0.90	0.95
	$\hat{p}_{EW}$	.354	.394	.484	.546	.605	.686	.720
10	$\hat{p}_{RT}$	.361	.401	.491	.552	.611	.692	.726
	$\hat{p}_{BT}$	.308	.344	.435	.505	.578	.686	.732
	$\hat{p}_{EW}$	.262	.314	.440	.531	.618	.733	.780
20	$\hat{p}_{RT}$	.264	.317	.443	.533	.620	.735	.781
	$\hat{p}_{BT}$	.236	.288	.417	.518	.615	.747	.798
	$\hat{p}_{EW}$	.204	.262	.412	.524	.632	.769	.822
30	$\hat{p}_{RT}$	.206	.264	.414	.523	.633	.770	.822
	$\hat{p}_{BT}$	.194	.249	.400	.518	.634	.776	.830
	$\hat{p}_{EW}$	.133	.193	.372	.516	.656	.822	.878
50	$\hat{p}_{RT}$	.133	.194	.374	.517	.656	.822	.878
	$\hat{p}_{BT}$	.128	.187	.361	.506	.653	.828	.885
	$\hat{p}_{EW}$	.052	.102	.308	.508	.703	.900	.948
100	$\hat{p}_{RT}$	.052	.102	.309	.508	.703	.899	.947
	$\hat{p}_{BT}$	.053	.101	.304	.506	.704	.903	.949

· $\hat{p}_{EW}$: The approximated value of the Edgeworth expansion
· $\hat{p}_{RT}$: The approximated value of the Roy–Tiku method
· $\hat{p}_{BT}$: The approximated value of the bootstrap method