Journal of the Korean Statistical Society (2006), **35***: 3, pp 331–341*

AN APPROXIMATE DISTRIBUTION OF THE SQUARED COEFFICIENT OF VARIATION UNDER GENERAL POPULATION†

YONGHEE LEE[1]

ABSTRACT

An approximate distribution of the plug-in estimator of the squared coefficient of variation (CV^2) is derived by using Edgeworth expansions under general population models. Also bias of the estimator is investigated for several important distributions. Under the normal distribution, we proposed the new estimator for CV^2 based on median of the sampling distribution of plug-in estimator.

AMS 2000 subject classifications. Primary 62E20; Secondary 62F12.
Keywords. Coefficient of variation, Edgeworth expansions, bootstrap, median.

1. INTRODUCTION

The squared coefficient of variation CV^2 represents the squared ratio of standard deviation to the mean for the given distribution

$$\mathrm{CV}^2 = \frac{\sigma^2}{\xi^2}\,,$$

where ξ and σ are the mean and the standard deviation, respectively, of a given distribution. It is the commonly used as a measure for comparing degree of variation with one distribution with another, even if the means of two distributions are different.

The squared coefficient of variation is used for the measure comparing inequality of certain indices across regions and times in economics. In fact, the

Received May 2006; accepted July 2006.
†This research was supported by KOSEF through Statistical Research Center for Complex Systems at Seoul National University.
[1]Department of Statistics, Ewha Womans University, Seoul 120-750, Korea (e-mail: ylee@ewha.ac.kr)

squared coefficient of variation is a special case of the class of generalized entropy measures for inequality (Cowell, 1995). Also in survey sampling and clinical trials, the formula for sample size determination is the function of CV^2 so that it is important to assess information about population CV^2 from available resources such as data from small pilot studies.

In practice, the population CV^2 is estimated by the simple plug-in estimator $\widehat{\mathrm{CV}^2}$ and it is a just squared ratio of sample standard deviation to sample mean

$$\widehat{\mathrm{CV}^2} = \frac{S^2}{\overline{X}^2}, \tag{1.1}$$

where $\overline{X}$ and S^2 are the sample mean and the sample variance, respectively, of the independent sample $X_1, X_2, \ldots, X_n$ from a given population distribution. Even though the estimator in (1.1) has its simple form, the sampling distribution of $\widehat{\mathrm{CV}^2}$ is not explicitly tractable. Even though $\widehat{\mathrm{CV}^2}$ has been extensively used for the sample size determination in clinical trials and survey sampling as well as for comparing inequality in economics, its sampling or asymptotic properties are not well developed. Recently, in econometrics, there have been some efforts to find the asymptotic properties of $\widehat{\mathrm{CV}^2}$. Breunig (2001) derived the asymptotic bias and mean squared error of $\widehat{\mathrm{CV}^2}$ and Giles (2005) used a small sigma expansion to derive the asymptotic bias of the general entropy measures including the squared coefficient of variation. In this paper, the approximation to the distribution of $\widehat{\mathrm{CV}^2}$ based on Edgeworth expansions is considered.

2. An Approximate Distribution of Estimator of CV^2 under General Populations

Edgeworth expansion is an asymptotic expansion to approximate the distribution of a given estimator $\hat{\theta}_n$ of unknown parameter θ when the limiting distribution of $\hat{\theta}_n$ is the normal distribution. Suppose that we are interested in approximating the distribution of the standardized statistics $U_n = \sqrt{n}(\hat{\theta}_n - \theta_0)/\tau$ where n is the sample size and τ^2 is the asymptotic variance of $\hat{\theta}_n$. Let $M_{U_n}(t) = E(e^{tU_n})$ be the moment generating function of U_n. Also let $K_{U_n}(t) = \log M_{U_n}(t)$ be the cumulant generating function of U_n and $\kappa_r(U_n)$ be the r^{th} cumulant of U_n.

Under some regularity conditions, the r^{th} cumulant of U_n is of the order of $n^{-(r-2)/2}$ and have an expansion with the power series in n^{-1} such that

$$\kappa_r(U_n) = n^{-(r-2)/2}\left[\kappa_{r1} + n^{-1}\kappa_{r2} + n^{-2}\kappa_{r3} + \cdots\right], \quad r \geq 1,$$

where $\kappa_{11} = 0$ and $\kappa_{21} = 1$. The Edgeworth expansion of distribution of U_n up to the second order is given by

$$P(U_n \leq t) = \Phi(t) + n^{-1/2} p_1(t)\phi(t) + O(n^{-1}) , \tag{2.1}$$

where $\Phi(t)$ is the distribution function of the standard normal distribution and

$$p_1(t) = -\left\{\kappa_{12} + \tfrac{1}{6}\kappa_{31}(t^2 - 1)\right\}$$

and $\phi(t)$ is the density of the standard normal distribution. The excellent review for Edgeworth expansions is given in Hall (1992).

Suppose $X_1, \ldots, X_n$ are independent sample from a given distribution with the population mean ξ and variance σ^2. We assume that the distribution is non-degenerate and absolutely continuous, which guarantees existence of Edgeworth expansion. The parameter of interest is the population squared coefficient of variation ($\theta = \sigma^2/\xi^2$). The sample squared coefficient of variation is defined as the commonly used plug-in estimator such that

$$\hat{\theta}_n = \frac{S^2}{\overline{X}^2} ,$$

where $\overline{X}$ is a sample mean and $S^2 = \sum(X_i - \overline{X})^2/n$ is a sample variance. Also we will assume that the population mean ξ is not zero for the proper definition of coefficient of variation. Let τ^2 be the asymptotic variance of $\hat{\theta}_n$ and define the p^{th} central moment of X as $m_p = E(X - \xi)^p$ and the standardized p^{th} central moment of X as $\gamma_p = E(X - \xi)^p/\sigma^p$.

PROPOSITION 2.1. *Under some regularity conditions, the Edgeworth expansion of distribution of* $U_n = \sqrt{n}(\hat{\theta}_n - \theta)/\tau$ *up to the second order is given by* (2.1) *with*

$$p_1(t) = -\left\{A_1\tau^{-1} + \tfrac{1}{6}A_2\tau^{-3}(t^2 - 1)\right\}$$

and

$$\tau^2 = 4\theta^3 + (\gamma_4 - 1)\theta^2 - 4\gamma_3\theta^{5/2}, \tag{2.2}$$

$$A_1 = 3\theta^2 - 2\gamma_3\theta^{3/2} - \theta, \tag{2.3}$$

$$\begin{aligned} A_2 = {} & \gamma_6\theta^3 - 6\gamma_5\theta^{7/2} + \gamma_4(36\theta^4 - 3\theta^3) - 12\gamma_4\gamma_3\theta^{7/2} + \gamma_3^2(42\theta^{8/2} - 6\theta^3) \\ & - \gamma_3(128\theta^{9/2} - 48\theta^{7/2} - 12\theta^{5/2}) + 72\theta^5 - 60\theta^4 + 2\theta^3, \end{aligned} \tag{2.4}$$

where A_1 *and* A_2 *are the first order term of the bias correction and skewness correction, respectively, in asymptotic expansion of cumulant of* U_n.

The regularity conditions and details of proof about Proposition 2.1 are given in Section 4.

REMARK 2.1. (Bias of $\widehat{\mathrm{CV}^2}$). Since $E[n^{1/2}(\hat{\theta}_n - \theta)] = n^{-1/2}A_1 + O(n^{-1})$ in Section 4, the bias of $\hat{\theta}_n$ has the asymptotic expansion such as

$$E(\hat{\theta}_n) - \theta = n^{-1}A_1 + O(n^{-3/2}) = n^{-1}(3\theta^2 - 2\gamma_3\theta^{3/2} - \theta) + O(n^{-3/2})$$

which implies that A_1 is the leading term in the asymptotic expansion of bias.

REMARK 2.2. (Normal distribution). If the population distribution is the normal distribution with mean ξ and variance σ^2, we have

$$\gamma_3 = \gamma_5 = 0, \quad \gamma_4 = 3 \quad \text{and} \quad \gamma_6 = 15.$$

Hence, the asymptotic variance of $\hat{\theta}_n$, A_1 and A_2 in (2.2), (2.3) and (2.4) can be simplified as

$$\tau^2 = 4\theta^3 + 2\theta^2, \quad A_1 = 3\theta^2 - \theta, \quad A_2 = 72\theta^5 + 48\theta^4 + 8\theta^3.$$

In particular, the bias of $\hat{\theta}_n$ has the asymptotic expansion

$$E(\hat{\theta}_n) - \theta = n^{-1}(3\theta^2 - \theta) + O(n^{-3/2}), \tag{2.5}$$

where the leading term is proportional to θ^2.

REMARK 2.3. (Gamma distribution). If the population distribution is gamma distribution with the density

$$f(x) = \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} \exp(-x/\beta),$$

the squared coefficient of variation is $\theta = \alpha^{-1}$ and the p^{th} cumulant of gamma distribution is given by $\kappa_p = (p-1)!\alpha\beta^p$. Using the formulae for the relation between cumulants and central moments in Kendall *et al.* (1987), we have

$$\tau^2 = 2\alpha^{-3} + 2\alpha^{-2}, \quad A_1 = -\alpha^{-2} - \alpha^{-1}, \quad A_2 = 32\alpha^{-5} + 40\alpha^{-4} + 32\alpha^{-3},$$

which also implies

$$\tau^2 = 2\theta^3 + 2\theta^2, \quad A_1 = -\theta^2 - \theta, \quad A_2 = 32\theta^5 + 40\theta^4 + 32\theta^3.$$

Since the sign of A_1 is negative, it can be known that the asymptotic bias of $\hat{\theta}_n$ is negative when gamma distribution is assumed.

REMARK 2.4. (Exponential distribution). The exponential distribution is a special case of gamma distribution with $\alpha = 1$ so that the squared coefficient of variation is a constant with $\theta = 1$ and it is easily shown that

$$\tau^2 = 4, \quad A_1 = -2, \quad A_2 = 104.$$

3. AN IMPROVED ESTIMATOR OF CV^2 UNDER THE NORMAL POPULATION

In the previous section, the asymptotic expansion of sampling distribution of $\hat{\theta}_n = \widehat{\mathrm{CV}^2}$ is derived under several important distributions by using Edgeworth expansion and also asymptotic expansion of bias is obtained. Under the normal distribution, it is shown that the bias of $\hat{\theta}_n$ increases as the order of θ^2 (see Remark 2.2). This implies the bias is proportional to ξ^{-2} as the population mean ξ converges to 0. For this reason in practice, use of $\hat{\theta}_n$ is restricted if the population mean is close to 0 under the normal population. Here, we derive the asymptotic expansion of median of the sampling distribution of $\hat{\theta}_n$ and propose the new estimator which has the smaller bias and mean squared error than $\hat{\theta}_n$.

The Fisher-Cornish expansion is an asymptotic expansion of the quantile of a given distribution and it is obtained by inverting the Edgeworth expansion of a given distribution. Details of Fisher-Cornish expansions is given in Hall (1983, 1992). Define x_α as the α-quantile of $U_n = n^{1/2}(\hat{\theta}_n - \theta)/\tau$ such as

$$P\{n^{1/2}(\hat{\theta}_n - \theta)/\tau \le x_\alpha\} = \alpha.$$

Then, by Fisher-Cornish expansion, α-quantile x_α has the following expansion

$$x_\alpha = z_\alpha - n^{-1/2}p_1(z_\alpha) + O(n^{-1}), \tag{3.1}$$

where $p_1(t)$ is the first leading term of the Edgeworth expansion of U_n in (2.1) and z_α is the α-quantile of standard normal distribution.

Let $\theta_{0.5}$ be the median of sampling distribution of $\hat{\theta}_n$ such that

$$P(\hat{\theta}_n \le \theta_{0.5}) = 0.5.$$

Then, by Fisher-Cornish expansion in (3.1) and since $z_{0.5} = 0$, $\theta_{0.5}$ has one-term expansion

$$\begin{aligned}\theta_{0.5} &= \theta + n^{-1/2}\tau x_{0.5} \\ &= \theta + n^{-1}(A_1 - \tfrac{1}{6}A_2\tau^{-2}) + o(n^{-1}).\end{aligned} \tag{3.2}$$

If the normal distribution is assumed as in Remark 2.2, the difference between $\theta_{0.5}$ and θ in the equation (3.2) has the following expansion

$$\theta_{0.5} - \theta = n^{-1}(A_1 - \tfrac{1}{6}A_2\tau^{-2}) + o(n^{-1})$$

$$
\begin{aligned}
&= n^{-1}\left(3\theta^2 - \theta - \frac{72\theta^5 + 48\theta^4 + 8\theta^3}{24\theta^3 + 12\theta^2}\right) + o(n^{-1}) \\
&= n^{-1}\left[3\theta^2 - \theta - \left\{3\theta^2 + \frac{\theta}{2} + \frac{1}{12} + O(\theta^{-1})\right\}\right] + o(n^{-1}) \\
&= n^{-1}\left\{-\frac{3}{2}\theta - \frac{1}{12} + O(\theta^{-1})\right\} + o(n^{-1}), \qquad (3.3)
\end{aligned}
$$

where the expansion in θ is valid as θ converges to infinity (*i.e.* the population mean converges to 0). While the leading bias term of $\hat{\theta}_n$ under normal distribution in (2.5) is proportional to θ^2, the leading term in expansion of $\theta_{0.5} - \theta$ is now proportional to θ. This fact implies if an estimator for $\theta_{0.5}$ is available, it may have less bias than $\hat{\theta}_n$ in estimating CV^2. Note that median $\theta_{0.5}$ is not the population median of the normal distribution which is same as the population mean ξ, but it is the median of sampling distribution of $\hat{\theta}_n$.

It is not easy to find an estimator of the median of $\hat{\theta}_n$ since the sampling distribution of $\hat{\theta}_n$ is not explicitly derived. However, we can use bootstrap to approximate the sampling distribution of $\hat{\theta}_n$ so that the estimator $\hat{\theta}_{0.5}^{boot}$ of $\theta_{0.5}$ can be obtained from the following equation based on the bootstrap distribution

$$P_*(\hat{\theta}_n^* \le \hat{\theta}_{0.5}^{boot} \mid X_1, \ldots, X_n) = 0.5,$$

where $P_*(\,\cdot\, \mid X_1, \ldots, X_n)$ is bootstrap distribution of $\hat{\theta}_n$ based on the sample $X_1, \ldots, X_n$ and $\hat{\theta}_n^*$ is the plug-in estimator of CV^2 based on the bootstrap sample $X_1^*, \ldots, X_n^*$.

Now, we propose a simple estimator $\hat{\theta}_{0.5}^{boot}$, which is denoted by $\widehat{\mathrm{CV}^2}_{BM}$, for the population CV^2 under the normal distribution based on median of the bootstrap distribution of $\hat{\theta}_n$. The new estimator $\widehat{\mathrm{CV}^2}_{BM}$ can be obtained by the following bootstrap method:

(1) Generates the h^{th} bootstrap sample $\mathcal{S}_h = \{X_1^*, \ldots, X_n^*\}$ from $\mathcal{S} = \{X_1, \ldots, X_n\}$ by a simple random sampling with replacement.

(2) Computes the ordinary estimate $\hat{\theta}_n^{*h} = \widehat{\mathrm{CV}^{2*}}$ based on the h^{th} bootstrap sample $\mathcal{S}_h = \{X_1^*, \ldots, X_n^*\}$.

(3) Repeats the steps (1) and (2) for a total of B times, which gives B values of $\hat{\theta}_n^{*h}$, $h = 1, \ldots, B$. Then, the median of $\hat{\theta}_n^{*h}$ provides the final estimate for CV^2, denoted by $\widehat{\mathrm{CV}^2}_{BM}$.

Note that the proposed estimator $\widehat{\mathrm{CV}^2}_{BM}$ is based on the asymptotic expansion of difference between the median of sampling distribution of $\hat{\theta}_n$ and the parameter CV^2 as given in (3.3). The idea used here is not a usual bias correction method which is based on the the asymptotic expansion of $E(\widehat{\mathrm{CV}^2})$. Hence, even though we may expect the proposed estimator may perform better than the naive estimator, the empirical performance of $\widehat{\mathrm{CV}^2}_{BM}$ should be investigated by Monte-Carlo simulation.

A small simulation study is conducted to compare performance of the proposed estimator $\widehat{\mathrm{CV}^2}_{BM}$ with the conventional estimator $\widehat{\mathrm{CV}^2}$. The sample sizes n =25, 50, 75 and 100 are considered under the normal distribution with $\xi/\sigma =$ 0.30, 0.40, 0.50 and 0.75, which implies that the values of parameter are $\mathrm{CV}^2 =$ 11.11, 6.25, 4.00 and 1.78. The bias and mean squared error (MSE) of two estimators are obtained by $M = 10,000$ Monte Carlo simulation and $B = 5,000$ bootstrap sample

$$\widehat{\text{bias}} = \frac{\sum_{m=1}^{M} \widehat{\mathrm{CV}^2}_{BM}(m)}{M} - \theta \quad \text{and} \quad \widehat{\text{MSE}} = \frac{\sum_{m=1}^{M} (\widehat{\mathrm{CV}^2}_{BM}(m) - \theta)^2}{M}.$$

Table 3.1 shows results of simulation study. If the value of CV^2 is less than 1.5, the bias and MSE of $\widehat{\mathrm{CV}^2}_{BM}$ and $\widehat{\mathrm{CV}^2}$ are almost identical so that they are omitted. As the value of population CV^2 becomes large, performance of conventional estimator $\widehat{\mathrm{CV}^2}$ is getting worse in sense that bias and MSE increase very rapidly. If sample size increases, performance of $\widehat{\mathrm{CV}^2}$ is getting better, but if CV^2 is larger than 10.0, increasing sample size does not improve bias and MSE. Regarding the proposed estimator $\widehat{\mathrm{CV}^2}_{BM}$, there is noticeable huge improvement both in bias and MSE over the plug-in estimator, especially when value of CV^2 is very large. Note that the bias of $\widehat{\mathrm{CV}^2}_{BM}$ is still positive even though the leading term in the asymptotic expansion of $\theta_{0.5} - \theta$ in (3.3) is negative. The reason is that the asymptotic expansion of $\theta_{0.5} - \theta$ is not same as that of $E(\widehat{\mathrm{CV}^2}) - \theta$ which is assessed by simulation study here. Also MSE of $\widehat{\mathrm{CV}^2}_{BM}$ shows some fluctuation when CV^2 is very large.

In conclusion, based on the simulation study, the proposed estimator is better than the conventional estimator $\widehat{\mathrm{CV}^2}$ when the population CV^2 is large and also two estimators show almost identical performance when the population CV^2 becomes small. Hence, it is recommended to use $\widehat{\mathrm{CV}^2}_{BM}$ if a underlying distribution is assumed to be the normal distribution.

TABLE 3.1 *Bias and mean squared error (MSE) of* $\widehat{CV^2}$ *and* $\widehat{CV^2}_{BM}$ *under the normal distribution*

			$\widehat{CV^2}$		$\widehat{CV^2}_{BM}$	
CV^2	ξ/σ	n	*Bias*	*MSE*	*Bias*	*MSE*
11.11	0.30	25	4.73×10^4	7.02×10^{12}	6.18	295.20
		50	1.41×10^3	3.12×10^9	9.36	678.90
		75	5.94×10^2	7.19×10^8	8.30	729.44
		100	4.73×10^1	2.46×10^6	7.29	631.61
6.25	0.40	25	5.45×10^2	2.28×10^8	4.55	175.86
		50	8.29×10^2	2.65×10^9	4.41	230.06
		75	5.54	1.66×10^4	2.63	125.32
		100	1.68	5.44×10^1	1.56	48.13
4.00	0.50	25	6.09×10^1	8.33×10^6	2.88	87.22
		50	6.24	7.63×10^4	1.41	43.99
		75	0.94	2.28×10^1	0.83	16.90
		100	0.62	8.27	0.55	7.83
1.77	0.75	25	0.74	9.15	0.45	6.34
		50	0.23	1.18	0.17	1.11
		75	0.15	0.60	0.11	0.57
		100	0.11	0.40	0.08	0.38

4. PROOF OF PROPOSITION 2.1

For notational convenience, the superscript of the vector indicates the corresponding component of vector, *i.e.* $\mathbf{x}^{(i)} = x^{(i)}$ is i^{th} component of the p-variate vector $\mathbf{x}'$ such that

$$\mathbf{x}' = (x^{(1)}, x^{(2)}, \ldots, x^{(i)}, \ldots, x^{(p)}).$$

Suppose $X_1, \ldots, X_n$ are independent sample from a given distribution with a population mean ξ and variance σ^2. Since the distribution of standardized quantity $S_n = n^{1/2}(\hat{\theta} - \theta)$ can be written as

$$\begin{aligned} P\{n^{1/2}(\hat{\theta} - \theta) \le x\} &= P\left\{ n^{1/2} \left(\frac{\sum X_i^2/n - \overline{X}^2}{\overline{X}^2} - \frac{\sigma^2}{\xi^2} \right) \le x \right\} \\ &= P\left\{ n^{1/2} \left(\frac{\sum X_i^2/n}{\overline{X}^2} - \frac{\sigma^2 + \xi^2}{\xi^2} \right) \le x \right\}, \end{aligned}$$

now we redefine the parameter of interest as $\theta = (\sigma^2 + \xi^2)/\xi^2$ and let its plug-in estimator be $\hat{\theta} = \sum X_i^2/(n\overline{X}^2)$. Also, we define the real-valued function f on R^2 such that $f(\mathbf{w}) = f(w^{(1)}, w^{(2)}) = w^{(2)}/(w^{(1)})^2$ and let $\mathbf{W} = (X, X^2)$,

$\overline{\mathbf{W}} = (\sum X_i/n, \sum X_i^2/n)$ and $\boldsymbol{\mu} = E(\mathbf{W}) = (\xi, \sigma^2 + \xi^2)$. Then, $\theta = f(\boldsymbol{\mu})$ and $\hat{\theta} = f(\overline{\mathbf{W}})$. Also put $\mathbf{Z} = n^{1/2}(\overline{\mathbf{W}} - \boldsymbol{\mu})$ and now by Taylor expansion we have the following expansion of $S_n = n^{1/2}(\hat{\theta} - \theta)$

$$n^{1/2}(\hat{\theta} - \theta) = \sum_{i=1}^{2} f_i Z^{(i)} + n^{-1/2}\frac{1}{2}\sum_{i=1}^{2}\sum_{j=1}^{2} f_{ij} Z^{(i)} Z^{(j)} + O_p(n^{-1}),$$

where $f_{i_1 i_2 \ldots i_p} = (\partial^p/\partial w^{(i_1)}\partial w^{(i_2)} \cdots \partial w^{(i_p)}) f(\mathbf{w})|_{\mathbf{w}=\boldsymbol{\mu}}$.

Put $\mu_{i_1 i_2 \ldots i_p} = E\{(\overline{\mathbf{W}} - \boldsymbol{\mu})^{(i_1)} \cdots (\overline{\mathbf{W}} - \boldsymbol{\mu})^{(i_p)}\}$ and define the p^{th} central moment of X as $m_p = E(X - \xi)^p$ and the standardized p^{th} central moment of X as $\gamma_p = E(X - \xi)^p/\sigma^p$. Then $\mu_i = 0$ for each i,

$$\begin{aligned} E(Z^{(i)} Z^{(j)}) &= \mu_{ij}, \\ E(Z^{(i)} Z^{(j)} Z^{(k)}) &= n^{-1/2}\mu_{ijk}, \\ E(Z^{(i)} Z^{(j)} Z^{(k)} Z^{(l)}) &= \mu_{ij}\mu_{kl} + \mu_{ik}\mu_{jl} + \mu_{il}\mu_{jk} + O(n^{-1}). \end{aligned}$$

Then,

$$\begin{aligned} E(S_n) &= n^{-1/2}\frac{1}{2}\sum_{i=1}^{2}\sum_{j=1}^{2} f_{ij}\mu_{ij} + O(n^{-1}), \\ E(S_n^2) &= \sum_{i=1}^{2}\sum_{j=1}^{2} f_i f_j \mu_{ij} + O(n^{-1}), \\ E(S_n^3) &= n^{-1/2}\Big\{\sum_{i=1}^{2}\sum_{j=1}^{2}\sum_{k=1}^{2} f_i f_j f_k \mu_{ijk} \\ &\quad + \frac{3}{2}\sum_{i=1}^{2}\sum_{j=1}^{2}\sum_{k=1}^{2}\sum_{l=1}^{2} f_i f_j f_{kl}(\mu_{ij}\mu_{kl} + \mu_{ik}\mu_{jl} + \mu_{il}\mu_{jk})\Big\} + O(n^{-1}). \end{aligned}$$

The asymptotic expansions of the three cumulants of S_n are

$$\begin{aligned} \kappa_{1,n} &= E(S_n) = n^{-1/2}A_1 + O(n^{-1}), \\ \kappa_{2,n} &= E(S_n^2) - \{E(S_n)\}^2 = \tau^2 + O(n^{-1}), \\ \kappa_{3,n} &= E(S_n^3) - 3E(S_n^2)E(S_n) + 2\{E(S_n)\}^3 = n^{-1/2}A_2 + O(n^{-1}), \end{aligned}$$

where

$$\tau^2 = \sum_{i=1}^{2}\sum_{j=1}^{2} f_i f_j \mu_{ij},$$

$$A_1 = \frac{1}{2}\sum_{i=1}^{2}\sum_{j=1}^{2} f_{ij}\mu_{ij},$$
$$A_2 = \sum_{i=1}^{2}\sum_{j=1}^{2}\sum_{k=1}^{2} f_i f_j f_k \mu_{ijk} + 3\sum_{i=1}^{2}\sum_{j=1}^{2}\sum_{k=1}^{2}\sum_{l=1}^{2} f_i f_j f_{kl}\mu_{ik}\mu_{jl}.$$

Hence, after some calculations, we have

$$\begin{aligned}
\mu_{11} &= m_2 \equiv \sigma^2,\\
\mu_{12} &= m_3 + 2\xi\sigma^2,\\
\mu_{22} &= m_4 + 4\xi m_3 + 4\xi^2\sigma^2 - \sigma^4,\\
\mu_{111} &= m_3,\\
\mu_{112} &= m_4 + 2\xi m_3 - \sigma^4,\\
\mu_{122} &= m_5 + 4\xi m_4 + (4\xi^2 - 2\sigma^2)m_3 - 4\xi\sigma^4,\\
\mu_{222} &= m_6 + 6\xi m_5 + (12\xi^2 - 3\sigma^2)m_4 + 8\xi^3 m_3 - 12\xi^2\sigma^4 + 2\sigma^6,
\end{aligned}$$

and $f_{22} = 0$, $f_1 = (-2)(\sigma^2 + \xi^2)\xi^{-3}$, $f_2 = \xi^{-2} f_{11} = 6(\sigma^2 + \xi^2)\xi^{-4}$, $f_{12} = f_{21} = (-2)\xi^{-3}$. Therefore, after some tedious algebraic calculations, we can obtain the asymptotic variance τ^2 of S_n as well as A_1 and A_2:

$$\begin{aligned}
\tau^2 &= \sum_{i=1}^{2}\sum_{j=1}^{2} f_i f_j \mu_{ij} = 4\theta^3 + (\gamma_4 - 1)\theta^2 - 4\gamma_3\theta^{5/2},\\
A_1 &= \frac{1}{2}\sum_{i=1}^{2}\sum_{j=1}^{2} f_{ij}\mu_{ij} = 3\theta^2 - 2\gamma_3\theta^{3/2} - \theta,\\
A_2 &= \sum_{i=1}^{2}\sum_{j=1}^{2}\sum_{k=1}^{2} f_i f_j f_k \mu_{ijk} + 3\sum_{i=1}^{2}\sum_{j=1}^{2}\sum_{k=1}^{2}\sum_{l=1}^{2} f_i f_j f_{kl}\mu_{ik}\mu_{jl}.\\
&= \gamma_6\theta^3 - 6\gamma_5\theta^{7/2} + \gamma_4(12\theta^4 - 3\theta^3) - \gamma_3(8\theta^{9/2} - 12\theta^{7/2} + 12\theta^{5/2}) - 12\theta^4 + 2\theta^3\\
&\quad +3\{8\gamma_4\theta^4 - 4\gamma_4\gamma_3\theta^{7/2} + \gamma_3^2(14\theta^{8/2} - 2\theta^3) - \gamma_3(40\theta^{9/2} - 12\theta^{7/2})\\
&\quad +24\theta^5 - 16\theta^4\}\\
&= \gamma_6\theta^3 - 6\gamma_5\theta^{7/2} + \gamma_4(36\theta^4 - 3\theta^3) - 12\gamma_4\gamma_3\theta^{7/2} + \gamma_3^2(42\theta^{8/2} - 6\theta^3)\\
&\quad -\gamma_3(128\theta^{9/2} - 48\theta^{7/2} - 12\theta^{5/2}) + 72\theta^5 - 60\theta^4 + 2\theta^3.
\end{aligned}$$

The regularity conditions for Edgeworth expansion are (i) existence of the finite moment such that $E(|X|^3) < \infty$, (ii) the differentiability of f, and (iii) the Cramer's condition which guarantees the distribution of X has a non-degenerate, absolutely continuous component. Under these regularity conditions, by using arguments in Hall (1992), we have one-term Edgeworth expansion of $S_n = n^{1/2}(\hat{\theta} - \theta)$ such that

$$P\{n^{1/2}(\hat{\theta} - \theta)/\tau \le x\} = \Phi(x) + n^{-1/2}p_1(x)\phi(x) + O(n^{-1}),$$

where

$$p_1(x) = -\left\{A_1\tau^{-1} + \tfrac{1}{6}A_2\tau^{-3}(x^2 - 1)\right\}.$$

REFERENCES

BREUNIG, R. (2001). "An almost unbiased estimator of the coefficient of variation", *Economics Letters*, **70**, 15–19.

COWELL, F. A. (1995). *Measuring Inequality*, 2nd ed., Prentice Hall/Harvester Wheatsheaf, London.

GILES, D. E. (2005). "The bias of inequality measures in very small samples: some analytic results", *Econometrics Working Papers*, **0514**.

HALL, P. (1983). "Inverting an Edgeworth expansion", *The Annals of Statistics*, **11**, 569–576.

HALL, P. (1992). *The Bootstrap and Edgeworth Expansion*, Springer-Verlag, New York.

KENDALL, M., STUART, A. AND ORD, J. K. (1987). *Kendall's Advanced Theory of Statistics*, Vol. 1, 5th ed., Oxford University Press, New York.