On the Distribution of Multivariate Sample Skewness for Assessing Multivariate Normality

Naoya Okamoto and Takashi Seo

Department of Mathematical Information Science, Faculty of Science,

Tokyo University of Science, 1-3, Kagurazaka, Shinjuku-ku, Tokyo 162-8601, Japan

Abstract

In this paper, we consider the multivariate normality test based on measure of multivariate sample skewness defined by Srivastava (1984). Srivastava derived asymptotic expectation up to the order N^{-1} for the multivariate sample skewness and approximate χ^2 test statistic, where N is sample size. Under normality, we derive the exact expectation and variance for Srivastava's multivariate sample skewness. From this result, improved approximate χ^2 test statistic using the multivariate sample skewness is also given for assessing multivariate normality. Finally, the numerical result by Monte Carlo simulation is shown in order to evaluate accuracy of the obtained expectation, variance and improved approximate χ^2 test statistic.

Key Words: asymptotic distribution; multivariate normal distribution; approximate χ^2 test statistic.

1. Introduction

In multivariate statistical analysis, the test for multivariate normality is an important problem and has been studied by many authors. Shapiro-Wilk (1965) derived test statistic using order statistic. Multivariate extensions of the Shapiro-Wilk (1965) test were proposed by Malkovich and Afifi (1973), Royston (1983), Srivastava and Hui (1987) and so on. Mardia (1970) and Srivastava (1984) gave different definitions of the multivariate sample skewness and kurtosis, and discussed test statistics using these measures for assessing multivariate normality. Mardia (1970) derived expectation of multivariate sample skewness and approximate χ^2 test statistic. Mardia (1974) derived more accurate approximate χ^2 test statistic using accessorial term than that of Mardia (1970). Test statistic using the multivariate sample kurtosis defined by Srivastava (1984) was discussed by Seo and Ariga (2006). Srivastava (1984) derived asymptotic expectation up to the order N^{-1} for multivariate sample skewness and approximate χ^2 test statistic using its asymptotic expectation. Thus, for small N, it seems that multivariate normality test using approximate χ^2 test statistic can not be carried out correctly.

In this paper, we consider the distribution of multivariate sample measure of skewness defined by Srivastava (1984). Exact expectation and variance of multivariate sample skewness are derived. Further, improved approximate χ^2 test statistic is also given by using exact expectation of multivariate sample skewness. Finally, we investigate accuracy of the exact expectation, the exact variance and distributions of approximate χ^2 test statistics derived by Srivastava (1984) and this paper by Monte Carlo simulation for selected parameters.

2. Distributions of multivariate sample skewness

2.1. Srivastava's multivariate sample skewness

Let $\boldsymbol{x}$ be a p-dimensional random vector with mean vector $\boldsymbol{\mu}$ and covariance matrix $\Sigma = \Gamma D_\lambda \Gamma'$, where $\Gamma = (\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \ldots, \boldsymbol{\gamma}_p)$ is an orthogonal matrix and $D_\lambda = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_p)$. Note that $\lambda_1, \lambda_2, \ldots, \lambda_p$ are the eigenvalues of Σ. Then, Srivastava (1984) defined the population measure of multivariate skewness as

$$\beta_{1,p}^2 = \frac{1}{p} \sum_{i=1}^{p} \left\{ \frac{\mathrm{E}[(y_i - \theta_i)^3]}{\lambda_i^{\frac{3}{2}}} \right\}^2 ,$$

where $y_i = \boldsymbol{\gamma}_i' \boldsymbol{x}$ and $\theta_i = \boldsymbol{\gamma}_i' \boldsymbol{\mu}$ $(i = 1, 2, \ldots, p)$. We note that $\beta_{1,p}^2 = 0$ under a multivariate normal population.

Let $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_N$ be samples of size N from a multivariate population. Let $\overline{\boldsymbol{x}}$ and $S = H D_\omega H'$ be sample mean vector and sample covariance matrix as follows:

$$\overline{\boldsymbol{x}} = \frac{1}{N} \sum_{j=1}^{N} \boldsymbol{x}_j,$$

$$S = \frac{1}{N} \sum_{j=1}^{N} (\boldsymbol{x}_j - \overline{\boldsymbol{x}})(\boldsymbol{x}_j - \overline{\boldsymbol{x}})',$$

respectively, where $H = (\boldsymbol{h}_1, \boldsymbol{h}_2, \ldots, \boldsymbol{h}_p)$ is an orthogonal matrix and $D_\omega = \mathrm{diag}(\omega_1, \omega_2, \ldots, \omega_p)$. We note that $\omega_1, \omega_2, \ldots, \omega_p$ are the eigenvalues of S. Then, Srivastava (1984) defined the sample measure of multivariate skewness as

$$b_{1,p}^2 = \frac{1}{p} \sum_{i=1}^{p} \left\{ \frac{1}{\omega_i^{\frac{3}{2}}} \sum_{j=1}^{N} \frac{(y_{ij} - \overline{y}_i)^3}{N} \right\}^2 , \qquad (1)$$

where $y_{ij} = \boldsymbol{h}_i'\boldsymbol{x}_j$ $(i = 1, 2, \ldots, p,\ j = 1, 2, \ldots, N)$, $\overline{y}_i = N^{-1}\sum_{j=1}^N y_{ij}$ $(i = 1, 2, \ldots, p)$.

For large N, Srivastava (1984) derived asymptotic expectation and approximate χ^2 test statistic for assessing multivariate normality as follows:

$$\mathrm{E}[b_{1,p}^2] = \frac{6}{N}, \tag{2}$$

$$\frac{Np}{6} b_{1,p}^2 \sim \chi_p^2, \tag{3}$$

respectively.

2.2. Exact expectation and variance for the multivariate sample skewness

In this subsection, we consider exact expectation $\mathrm{E}[b_{1,p}^2]$ and exact variance $\mathrm{Var}[b_{1,p}^2]$ for Srivastava's multivariate sample skewness $b_{1,p}^2$ in (1) under normality. In order to avoid the dependence of y_{ij} and $\overline{y}_i$, let $\overline{y}_i^{(\alpha)}$ be a mean defined on the subset of $y_{i1}, y_{i2}, \ldots, y_{iN}$, that is,

$$\overline{y}_i^{(\alpha)} = \frac{1}{N-1} \sum_{j=1,\, j\neq\alpha}^{N} y_{ij}.$$

Note that $y_{i\alpha}$ is independent of $\overline{y}_i^{(\alpha)}$. In order to obtain the expectation of $b_{1,p}^2$, we put

$$\overline{y}_i^{(\alpha)} = \frac{1}{\sqrt{N-1}} Z_i.$$

Since Z_i is distributed as standard normal distribution, the odd order moments equal zero and

$$\mathrm{E}[Z_i^{2k}] = (2k-1)\cdots 5 \cdot 3 \cdot 1, \quad k = 1, 2, \ldots, 6.$$

We note that

$$\begin{aligned} \omega_i &= \boldsymbol{h}_i' S \boldsymbol{h}_i \\ &= \frac{1}{N} \sum_{j=1}^{N} (y_{ij} - \overline{y}_i)^2, \quad i = 1, 2, \ldots, p. \end{aligned}$$

Then, we can write

$$\begin{aligned} b_{1,p}^2 &= \frac{1}{p} \sum_{i=1}^{p} \left\{ \frac{m_{3i}}{m_{2i}^{\frac{3}{2}}} \right\}^2 \\ &= \frac{1}{p} \sum_{i=1}^{p} b_{1(i)}^2, \end{aligned}$$

where $b_{1(i)} = m_{3i}/m_{2i}^{3/2}$, $m_{\nu i} = N^{-1}\sum_{j=1}^{N}(y_{ij} - \overline{y}_i)^{\nu}$. Note that $\boldsymbol{h}_i \to \boldsymbol{\gamma}_i$ with probability one. Hence, under the hypothesis of normality, for large N, we get

$$y_{ij} = \boldsymbol{h}_i'\boldsymbol{x}_j \sim N(\boldsymbol{\gamma}_i'\boldsymbol{\mu}, \lambda_i).$$

Thus, $y_{i1}, y_{i2}, \ldots, y_{iN}$ are asymptotically independently normally distributed. Hence, for large N,

$$\mathrm{E}[m_{\nu i}^{k} m_{2i}^{-\nu k/2}]\mathrm{E}[m_{2i}^{\nu k/2}] = \mathrm{E}[m_{\nu i}^{k}].$$

Therefore, we can treat each expectation of numerator and denominator of $b_{1(i)}^2$.

Now consider $\mathrm{E}[b_{1(i)}^2] = \mathrm{E}[m_{3i}^2]/\mathrm{E}[m_{2i}^3]$, where

$$\begin{aligned}
\mathrm{E}[m_{3i}^2] &= \frac{1}{N^2}\left\{N\mathrm{E}[C_{i\alpha}^6] + N(N-1)\mathrm{E}[C_{i\alpha}^3 C_{i\beta}^3]\right\}, \\
\mathrm{E}[m_{2i}^3] &= \frac{1}{N^3}\{N\mathrm{E}[C_{i\alpha}^6] + 3N(N-1)\mathrm{E}[C_{i\alpha}^4 C_{i\beta}^2] + N(N-1)(N-2)\mathrm{E}[C_{i\alpha}^2 C_{i\beta}^2 C_{i\gamma}^2]\},
\end{aligned}$$

and $C_{i\alpha} = y_{i\alpha} - \overline{y}_i$. If needed, let $\overline{y}_i^{(\alpha,\beta)}$ be a mean defined on the subset of $y_{i1}, y_{i2}, \ldots, y_{iN}$ by deleting variables $y_{i\alpha}$ and $y_{i\beta}$, that is,

$$\overline{y}_i^{(\alpha,\beta)} = \frac{1}{N-2}\sum_{j=1,\, j\neq\alpha,\beta}^{N} y_{ij},$$

and let

$$\overline{y}_i^{(\alpha,\beta)} = \frac{1}{\sqrt{N-2}}W_i,$$

where W_i is distributed as standard normal distribution.

After a great deal of calculation of expectations, we obtain

$$\begin{aligned}
\mathrm{E}[m_{3i}^2] &= \frac{6}{N} - \frac{18}{N^2} + \frac{12}{N^3}, \\
\mathrm{E}[m_{2i}^3] &= 1 + \frac{3}{N} - \frac{1}{N^2} - \frac{3}{N^3},
\end{aligned}$$

and we can obtain the expectation for $b_{1(i)}^2$ as

$$\begin{aligned}
\mathrm{E}[b_{1(i)}^2] &= \left(\frac{6}{N} - \frac{18}{N^2} + \frac{12}{N^3}\right) \Big/ \left(1 + \frac{3}{N} - \frac{1}{N^2} - \frac{3}{N^3}\right) \\
&= \frac{6(N-2)}{(N+1)(N+3)}.
\end{aligned}$$

Therefore, we have the expectation for $b_{1,p}^2$ as

$$\begin{aligned}
\mathrm{E}[b_{1,p}^2] &= \frac{1}{p}\sum_{i=1}^{p}\mathrm{E}[b_{1(i)}^2] \\
&= \frac{6(N-2)}{(N+1)(N+3)}.
\end{aligned}$$

As for the variances for $b_{1(i)}^2$ and $b_{1,p}^2$, we have

$$\begin{aligned}
\mathrm{Var}[b_{1(i)}^2] &= \mathrm{E}[(b_{1(i)}^2)^2] - \left\{\mathrm{E}[b_{1(i)}^2]\right\}^2, \\
\mathrm{Var}[b_{1,p}^2] &= \frac{1}{p^2}\sum_{i=1}^{p} \mathrm{Var}[b_{1(i)}^2] \\
&= \frac{1}{p^2}\sum_{i=1}^{p}\left(\frac{\mathrm{E}[m_{3i}^4]}{\mathrm{E}[m_{2i}^6]} - \left\{\mathrm{E}[b_{1(i)}^2]\right\}^2\right),
\end{aligned}$$

where

$$\begin{aligned}
\mathrm{E}[m_{3i}^4] \;=\; & \frac{1}{N^4}\left\{N\mathrm{E}[C_{i\alpha}^{12}] + 4N(N-1)\mathrm{E}[C_{i\alpha}^9 C_{i\beta}^3] + 3N(N-1)\mathrm{E}[C_{i\alpha}^6 C_{i\beta}^6]\right. \\
& +6N(N-1)(N-2)\mathrm{E}[C_{i\alpha}^6 C_{i\beta}^3 C_{i\gamma}^3] \\
& \left.+N(N-1)(N-2)(N-3)\mathrm{E}[C_{i\alpha}^3 C_{i\beta}^3 C_{i\gamma}^3 C_{i\delta}^3]\right\}, \\
\mathrm{E}[m_{2i}^6] \;=\; & \frac{1}{N^6}\left\{N\mathrm{E}[C_{i\alpha}^{12}] + 6N(N-1)\mathrm{E}[C_{i\alpha}^{10} C_{i\beta}^2] + 15N(N-1)\mathrm{E}[C_{i\alpha}^8 C_{i\beta}^4]\right. \\
& +15N(N-1)(N-2)\mathrm{E}[C_{i\alpha}^8 C_{i\beta}^2 C_{i\gamma}^2] + 10N(N-1)\mathrm{E}[C_{i\alpha}^6 C_{i\beta}^6] \\
& +60N(N-1)(N-2)\mathrm{E}[C_{i\alpha}^6 C_{i\beta}^4 C_{i\gamma}^2] \\
& +20N(N-1)(N-2)(N-3)\mathrm{E}[C_{i\alpha}^6 C_{i\beta}^2 C_{i\gamma}^2 C_{i\delta}^2] \\
& +15N(N-1)(N-2)\mathrm{E}[C_{i\alpha}^4 C_{i\beta}^4 C_{i\gamma}^4] \\
& +45N(N-1)(N-2)(N-3)\mathrm{E}[C_{i\alpha}^4 C_{i\beta}^4 C_{i\gamma}^2 C_{i\delta}^2] \\
& +15N(N-1)(N-2)(N-3)(N-4)\mathrm{E}[C_{i\alpha}^4 C_{i\beta}^2 C_{i\gamma}^2 C_{i\delta}^2 C_{i\varepsilon}^2] \\
& \left.+N(N-1)(N-2)(N-3)(N-4)(N-5)\mathrm{E}[C_{i\alpha}^2 C_{i\beta}^2 C_{i\gamma}^2 C_{i\delta}^2 C_{i\varepsilon}^2 C_{i\xi}^2]\right\}.
\end{aligned}$$

Note that $b_{1(i)}$ and $b_{1(j)}$ $(i \neq j)$ are independent.

After a great deal of calculation by using the same way as mentioned above, we obtain

$$\begin{aligned}
\mathrm{E}[m_{3i}^4] &= \frac{108}{N^2} + \frac{2592}{N^3} - \frac{16092}{N^4} + \frac{28512}{N^5} - \frac{15120}{N^6}, \\
\mathrm{E}[m_{2i}^6] &= 1 + \frac{24}{N} + \frac{205}{N^2} + \frac{720}{N^3} + \frac{739}{N^4} - \frac{744}{N^5} - \frac{945}{N^6},
\end{aligned}$$

and we can obtain the variances for $b_{1(i)}^2$ and $b_{1,p}^2$ as

$$\begin{aligned}
\mathrm{Var}[b_{1(i)}^2] &= \frac{\left(\frac{108}{N^2} + \frac{2592}{N^3} - \frac{16092}{N^4} + \frac{28512}{N^5} - \frac{15120}{N^6}\right)}{\left(1 + \frac{24}{N} + \frac{205}{N^2} + \frac{720}{N^3} + \frac{739}{N^4} - \frac{744}{N^5} - \frac{945}{N^6}\right)} - \left\{\frac{6(N-2)}{(N+1)(N+3)}\right\}^2 \\
&= \frac{72N(N-2)(N^3+37N^2+11N-313)}{(N+1)^2(N+3)^2(N+5)(N+7)(N+9)}, \\
\mathrm{Var}[b_{1,p}^2] &= \frac{1}{p}\frac{72N(N-2)(N^3+37N^2+11N-313)}{(N+1)^2(N+3)^2(N+5)(N+7)(N+9)},
\end{aligned}$$

respectively.

Therefore, we have the following theorem on exact expectation and variance for $b_{1,p}^2$.

Theorem 1 *The exact expectation and variance for multivariate sample skewness $b_{1,p}^2$ are given by*

$$\mathrm{E}[b_{1,p}^2] = \frac{6(N-2)}{(N+1)(N+3)}, \tag{4}$$

$$\mathrm{Var}[b_{1,p}^2] = \frac{1}{p}\frac{72N(N-2)(N^3+37N^2+11N-313)}{(N+1)^2(N+3)^2(N+5)(N+7)(N+9)}, \tag{5}$$

respectively.

2.3. Improved approximate χ^2 test statistic

In this subsection, we discuss improved test statistic of approximate χ^2 test statistic (3) by Srivastava (1984) for assessing multivariate normality.

Let $\mathrm{E}[b_{1,p}^2] = s^2$, then, for large N, we can write $b_{1(i)} \sim N(0, s^2)$, and $(1/s)b_{1(i)} \sim N(0,1)$. Then, we have

$$\frac{1}{s^2}\sum_{i=1}^{p} b_{1(i)}^2 = \frac{1}{s^2} p b_{1,p}^2 \sim \chi_p^2.$$

Therefore, we can obtain improved χ^2 test statistic as

$$\frac{(N+1)(N+3)}{6(N-2)} p b_{1,p}^2 \sim \chi_p^2. \tag{6}$$

3. Accuracy of distributions of $b_{1,p}^2$ and χ^2 test statistic

Accuracy of exact expectation, exact variance and percentiles of approximate χ^2 distribution is evaluated via a Monte Carlo simulation study for the following parameters: $p = 3, 5, 7, 10$, $N = 20, 50, 100, 200, 400, 800$. As a numerical experiment, we carry out 10,000,000 replications.

Table 1 gives Srivastava's result (expectation), exact values (expectation and variance) derived by this paper and simulated values (expectation and variance) derived by Monte Carlo simulation. In expectation of Table 1, "*Srivastava*" is values calculated by using (2) and "*Ours*" is values calculated by using (4). In variance of Table 1, "*Ours*" is values calculated by using (5). Since Srivastava did not derive variance of multivariate sample skewness, there is not that in Table 1. Figures 1–2 show the graphs of expectation of the corresponding data in Table 1 for $p = 3$, 10. In Table 1 and Figures 1–2, *Ours* is closer to simulated values than *Srivastava* for all parameters. *Ours* and simulated values are almost same even when N is small. On the other hand, we note that *Srivastava* is not close to simulated values when N is small. For variance of Table 1, *Ours* and simulated values are almost same. For expectation and variance of Table 1, we confirmed *Ours* and simulated values are almost same.

Tables 2–3 give values of the upper 5, 2.5 percentiles and lower 5, 2.5 percentiles of the distribution of approximate χ^2 test statistic derived by Srivastava (1984) and this paper, respectively. In Tables 2–3, "*Srivastava*" is values calculated by using (3) and "*Ours*" is values calculated by using (6). Figures 3–4 show the graphs of the corresponding data in Table 2, and Figures 5–6 show those in Table 3 for $p = 3,\ 10$. Table 2 and Figures 3–4 show that *Ours* is closer to upper percentiles of χ^2_p than *Srivastava* when N is small. When N is large, *Ours* is close to percentiles of χ^2_p as well as *Srivastava*. Table 3 and Figures 5–6 show that *Ours* is closer to lower percentiles of χ^2_p than *Srivastava* for all dimensions and sample sizes.

Table 4 gives expectations and variances of test statistics for multivariate normality test by Srivastava (1984) and this paper. In Table 4, "*Srivastava*" is values calculated by using (3) and "*Ours*" is values calculated by using (6). For expectation, *Ours* is closer to expectation of χ^2 distribution with p degrees of freedom $\mathrm{E}(\chi^2_p)(= p)$ than *Srivastava* for all dimensions and sample sizes even when N is small. For variance, we confirmed *Ours* is close to $\mathrm{Var}(\chi^2_p)(= 2p)$ as well as *Srivastava* when N is large.

Figures 7–8 show the graphs of probability density function of χ^2 distribution with p degrees of freedom, and frequency distributions of (3) by *Srivastava* and (6) by this paper (*Ours*) based on 10,000,000 values of test statistics for $p = 3$, $N = 20$ and $p = 10$, $N = 20$, respectively. These figures show that the frequency distribution of *Ours* is closer to the probability density function of χ^2_p than that of *Srivastava* even when N is small. On the other hand, *Srivastava* shows gap to the left in relation from probability density function of χ^2_p and expectation of *Srivastava* is smaller than the value of p.

In conclusion, it may be noted from these simulation results that the improved χ^2 test statistic proposed in this paper is useful for the multivariate normality test.

References

Malkovich, J. F. and Afifi, A. A. (1973). On tests for multivariate normality. *J. Amer. Statist. Assoc.*, **68**, 176–179.

Mardia, K. V. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika*, **57**, 519–530.

Mardia, K. V. (1974). Applications of some measures of multivariate skewness and kurtosis in testing normality and robustness studies. *Sankhyá*, series B: **36**, 115–128.

Royston, J. P. (1983). Some techniques for assessing multivariate normality based on the Shapiro-Wilk W. *Appl. Statist.*, **32**, 121–133.

Seo, T. and Ariga, M. (2006). On the distribution of kurtosis test for multivariate normality. Technical Report No.06-03. Statistical Research Group, Hiroshima University.

Shapiro, S. S. and Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, **52**, 591–611.

Srivastava, M. S. (1984). A measure of skewness and kurtosis and a graphical method for assessing multivariate normality. *Statist.* & *Prob. letters*, **2**, 263–267.

Srivastava, M. S. and Hui, T. K. (1987). On assessing multivariate normality based on Shapiro-Wilk W statistic. *Statist.* & *Prob. letters*, **5**, 15–18.

Table 1: Expectation and variance of multivariate sample skewness $b_{1,p}^2$.

		Expectation			Variance	
p	N	*Srivastava*	*Ours*	simulation	*Ours*	simulation
3	20	0.3000	0.2236	0.2237	0.0430	0.0430
	50	0.1200	0.1065	0.1066	0.0093	0.0093
	100	0.0600	0.0565	0.0565	0.0024	0.0024
	200	0.0300	0.0291	0.0291	0.0006	0.0006
	400	0.0150	0.0148	0.0148	0.0002	0.0002
	800	0.0075	0.0074	0.0074	0.0000	0.0000
5	20	0.3000	0.2236	0.2237	0.0258	0.0258
	50	0.1200	0.1065	0.1066	0.0056	0.0056
	100	0.0600	0.0565	0.0565	0.0015	0.0015
	200	0.0300	0.0291	0.0291	0.0004	0.0004
	400	0.0150	0.0148	0.0148	0.0001	0.0001
	800	0.0075	0.0074	0.0074	0.0000	0.0000
7	20	0.3000	0.2236	0.2236	0.0184	0.0184
	50	0.1200	0.1065	0.1066	0.0040	0.0040
	100	0.0600	0.0565	0.0565	0.0010	0.0010
	200	0.0300	0.0291	0.0291	0.0003	0.0003
	400	0.0150	0.0148	0.0148	0.0001	0.0001
	800	0.0075	0.0074	0.0074	0.0000	0.0000
10	20	0.3000	0.2236	0.2236	0.0129	0.0129
	50	0.1200	0.1065	0.1065	0.0028	0.0028
	100	0.0600	0.0565	0.0565	0.0007	0.0007
	200	0.0300	0.0291	0.0291	0.0002	0.0002
	400	0.0150	0.0148	0.0148	0.0000	0.0000
	800	0.0075	0.0074	0.0074	0.0000	0.0000

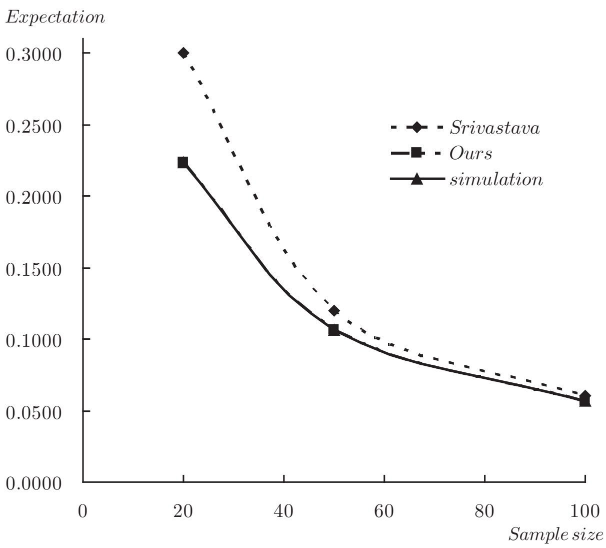

Figure 1: Expectation of $b_{1,p}^2$ for $p = 3$.

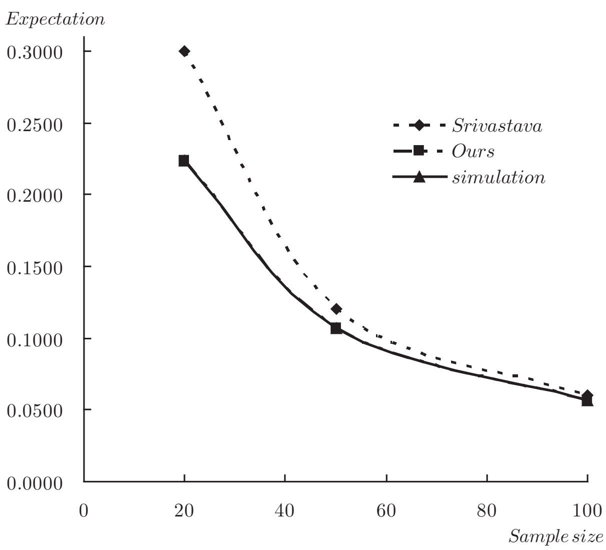

Figure 2: Expectation of $b_{1,p}^2$ for $p = 10$.

Table 2: Upper 5 and 2.5 percentiles of approximate χ^2 test statistic.

		Upper 5 percentile			Upper 2.5 percentile		
p	N	*Srivastava*	*Ours*	$\chi_p^2(0.05)$	*Srivastava*	*Ours*	$\chi_p^2(0.025)$
3	20	6.24	8.12	7.81	7.77	10.10	9.35
	50	7.29	8.16		9.00	10.08	
	100	7.60	8.06		9.29	9.85	
	200	7.73	7.96		9.37	9.65	
	400	7.78	7.90		9.37	9.51	
	800	7.80	7.86		9.36	9.43	
5	20	8.52	11.72	11.07	10.20	14.03	12.83
	50	10.20	11.44		12.14	13.62	
	100	10.76	11.41		12.69	13.45	
	200	10.92	11.25		12.80	13.18	
	400	11.01	11.18		12.84	13.04	
	800	11.05	11.13		12.84	12.93	
7	20	11.10	14.96	14.07	12.99	17.51	16.01
	50	13.10	14.69		15.28	17.12	
	100	13.64	14.46		15.77	16.72	
	200	13.91	14.33		15.98	16.46	
	400	13.99	14.20		16.00	16.24	
	800	14.04	14.15		16.01	16.13	
10	20	14.45	19.27	18.31	16.56	22.08	20.48
	50	17.04	19.09		19.47	21.81	
	100	17.77	18.84		20.14	21.35	
	200	18.10	18.65		20.41	21.02	
	400	18.21	18.49		20.46	20.77	
	800	18.26	18.40		20.47	20.63	

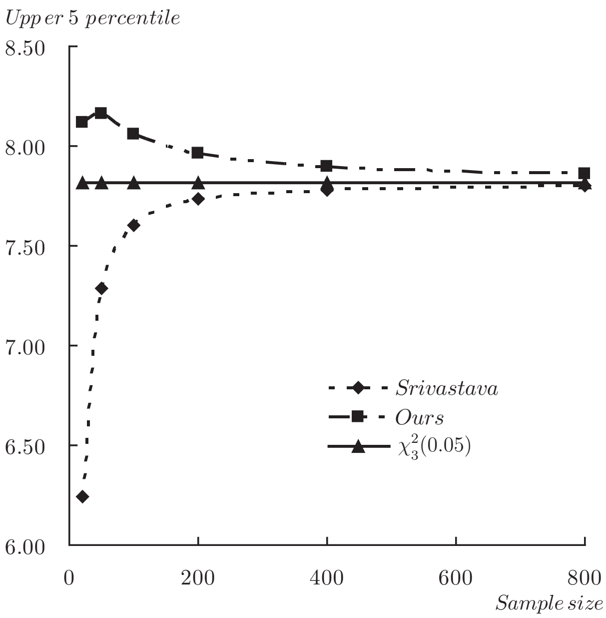

Figure 3: Upper 5 percentile for $p = 3$.

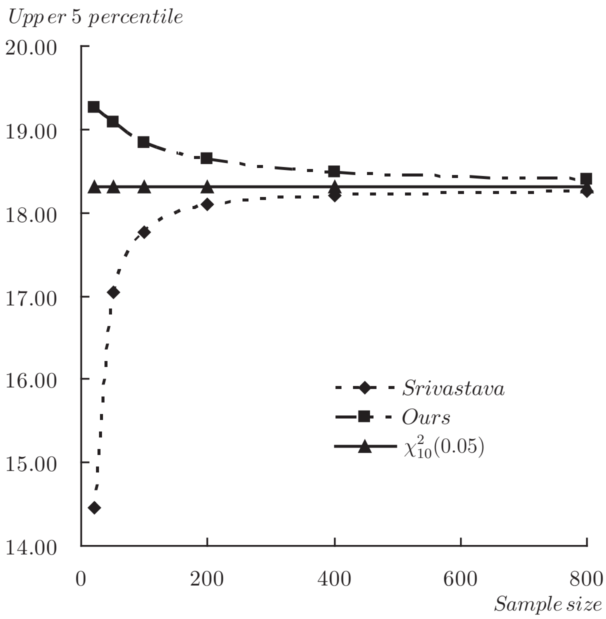

Figure 4: Upper 5 percentile for $p = 10$.

Table 3: Lower 5 and 2.5 percentiles of approximate χ^2 test statistic.

		Lower 5 percentile			Lower 2.5 percentile		
p	N	*Srivastava*	*Ours*	$\chi^2_p(0.95)$	*Srivastava*	*Ours*	$\chi^2_p(0.975)$
3	20	0.23	0.30	0.35	0.14	0.19	0.22
	50	0.29	0.32		0.18	0.20	
	100	0.31	0.33		0.19	0.20	
	200	0.33	0.34		0.20	0.21	
	400	0.34	0.35		0.21	0.21	
	800	0.35	0.35		0.21	0.21	
5	20	0.74	1.01	1.15	0.53	0.73	0.83
	50	0.93	1.04		0.67	0.75	
	100	1.02	1.08		0.74	0.78	
	200	1.07	1.11		0.78	0.80	
	400	1.11	1.13		0.80	0.82	
	800	1.13	1.13		0.82	0.82	
7	20	1.44	1.94	2.17	1.12	1.51	1.69
	50	1.78	2.00		1.38	1.55	
	100	1.94	2.05		1.51	1.60	
	200	2.04	2.10		1.59	1.64	
	400	2.10	2.13		1.64	1.66	
	800	2.13	2.15		1.66	1.67	
10	20	2.66	3.55	3.94	2.18	2.90	3.25
	50	3.27	3.66		2.68	3.00	
	100	3.55	3.76		2.91	3.09	
	200	3.73	3.84		3.07	3.16	
	400	3.83	3.88		3.15	3.20	
	800	3.88	3.91		3.20	3.22	

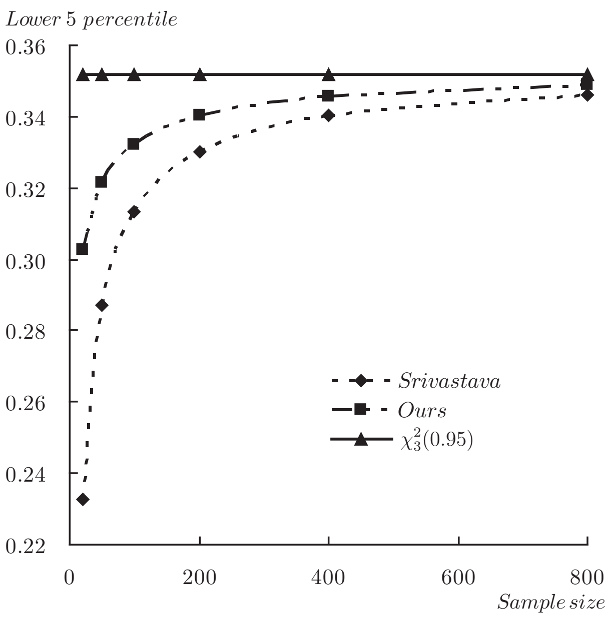

Figure 5: Lower 5 percentile for $p = 3$.

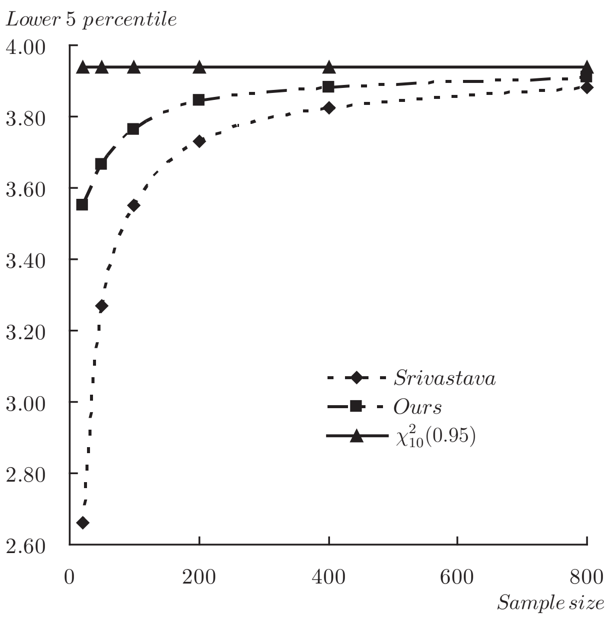

Figure 6: Lower 5 percentile for $p = 10$.

Table 4: Expectation and variance of approximate χ^2 test statistic.

		Expectation			Variance		
p	N	*Srivastava*	*Ours*	$\mathrm{E}(\chi^2_p)$	*Srivastava*	*Ours*	$\mathrm{Var}(\chi^2_p)$
3	20	2.24	3.00	3.00	4.30	7.75	6.00
	50	2.66	3.00		5.80	7.36	
	100	2.83	3.00		6.09	6.86	
	200	2.91	3.00		6.10	6.48	
	400	2.96	3.00		6.08	6.26	
	800	2.98	3.00		6.08	6.17	
5	20	3.73	5.00	5.00	7.16	12.90	10.00
	50	4.44	5.00		9.67	12.26	
	100	4.71	5.00		10.13	11.42	
	200	4.85	5.00		10.17	10.79	
	400	4.93	5.00		10.11	10.42	
	800	4.96	5.00		10.22	10.38	
7	20	5.22	7.00	7.00	10.02	18.04	14.00
	50	6.22	7.00		13.55	17.18	
	100	6.59	7.00		14.20	16.00	
	200	6.79	7.00		14.21	15.09	
	400	6.89	7.00		14.16	14.59	
	800	6.95	7.00		13.94	14.15	
10	20	7.45	10.00	10.00	14.32	25.78	20.00
	50	8.88	10.00		19.34	24.53	
	100	9.42	10.00		20.28	22.85	
	200	9.71	10.00		20.33	21.59	
	400	9.85	10.00		20.44	21.07	
	800	9.93	10.00		19.56	19.85	

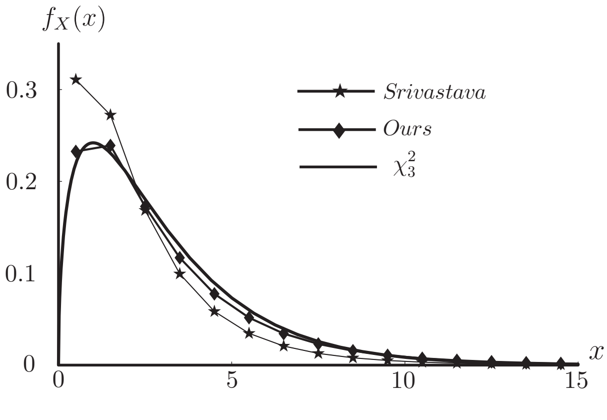

Figure 7: Distribution for $p = 3$, $N = 20$.

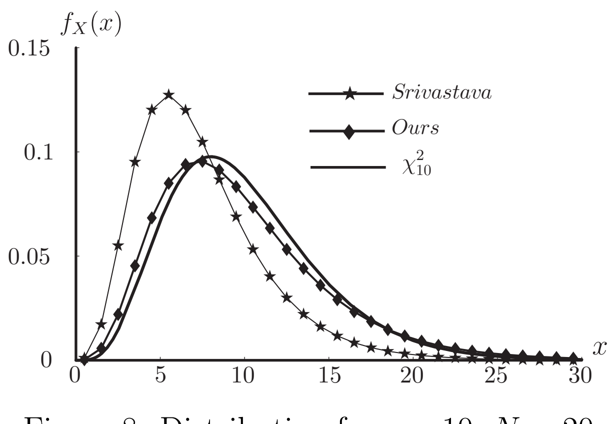

Figure 8: Distribution for $p = 10$, $N = 20$.