▩ 연구논문

A Study of Control Chart for Skewness[+]

Jung Jin Lee
Dept. of Statistics, Soong Sil University

Abstract

Sample skewness has not received much attention from researchers to design a control chart. In this paper, control charts based on two skewness measures are studied to control a manufacturing process. One skewness measure is the third central moment about mean, the other is the third L-moment which is a linear combination of order statistics. Since the exact sampling distributions of two skewness measures are unknown, empirical sampling distributions are studied using simulation. The sampling distributions are used to design control charts for skewness and performance of two skewness measures is compared.

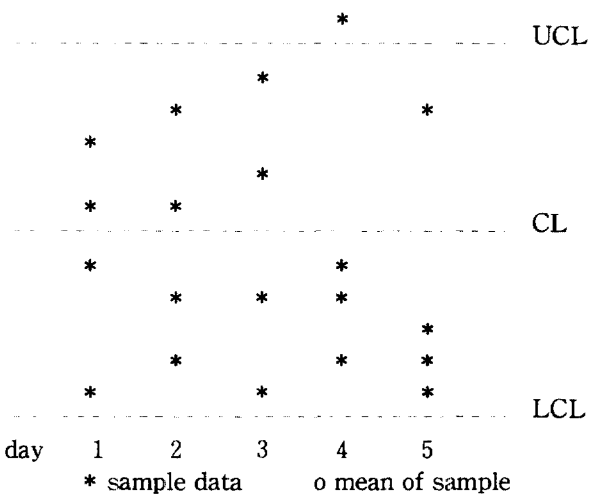
1. Introduction

The existing control charts for variable utilize the sample statistics of the measurements from a manufacturing process. The control limits of the control charts are determined using sampling distributions of sample statistics. When a process is under control, almost all sample statistics fall within their control limits. If a sample statistic falls outside the control limits, we investigate whether the process is out of control. Also, even if the sample statistics are within the control limits, several statistical tests are used to check whether the process is under control.

The sample statistics used in the control charts are sample mean, range, standard deviation etc. Sample skewness has not received much attention from researchers to design a control chart [Duncan, 1974]. One of the reasons is that the

+ This research was supported in part by Soong Sil University.

exact sampling distribution of skewness is unknown. However, the $\overline{X}$ chart in ⟨Figure 1⟩ shows that sample skewness may also be used to check whether a process is under control or not. All sample data are under control based on the $\overline{X}$ chart, but the process seems to be out of control on day 4 and 5 and needs investigation.



⟨ Figure 1 ⟩ An example of $\overline{X}$ chart whose skewness of data is out of control

In this paper, a control chart of skewness is discussed to control the process. The traditional measure of skewness is the third central moment about mean. Recently, Hosking proposed L-skewness based on the linear combination of order statistics which are more robust than conventional moments to outliers in the data. In section 2, the definitions and properties of two skewness measures are reviewed. In section 3, sampling distributions of two skewness measures in case of small sample are studied using simulation. In section 4, the control charts of two skewness measures are compared by various numerical experiments. In section 5, the conclusion of this paper is discussed.

2. Skewness Measures

The most commonly used measure of skewness is the third central moment about mean. Recently, the third L-moment which uses L_1 norm of order statistic is proposed by Hosking (1990) as a measure of skewness. The definitions and properties of the two skewness measures are reviewed in this section.

2.1 Third Central Moment

Let X be a real valued random variable with mean μ and variance σ^2 and distribution function $F(x)$. The traditional measure of skewness, σ_3, is the third central moment about mean defined as following:

$$\sigma_3 = E[(\frac{X-\mu}{\sigma})^3]$$

If the value of σ_3 is positive(negative), then the distribution of X is skewed to the right(left) from the mean. If the value is zero, then the distribution is symmetric about mean. Since this measure uses a L_3 norm, it is sensitive to an extreme data.

Let $X_1, X_2, \ldots, X_n$ be a random sample of size n drawn from the distribution of X and let $\overline{X}$ and s^2 be the sample mean and variance respectively. The third central moment of sample, s_3, is defined as following:

$$s_3 = \frac{1}{n}\sum_i(\frac{X_i - X}{s})^3$$

Note that s_3 is invariant under linear transformation of X_i. The exact distribution of s_3 is unknown. However, if the sample size is sufficiently large and the population distribution is normal with mean μ and variance σ^2, then s_3 is asymptotically normal with mean 0 and variance as follows [Wilks, 1962]:

$$Var(s_3) = \frac{6(n-2)}{(n+1)(n+3)}$$

2.2 Third L-moment

Let X be a real-valued random variable with distribution function $F(x)$ and quantile function $x(F)$, and let $X_{1:n} \le X_{2:n} \le \cdots \le X_{n:n}$ be the order statistics of a random sample of size n drawn from the distribution of X. Define the L-moments of X to be the quantities

$$\lambda_r = r^{-1}\sum_{k=0}^{r-1}(-1)^k \binom{r-1}{k} EX_{r-k:r}, \qquad r=1, 2, \cdots$$

The L in 'L-moments' emphasizes that λ_r is a linear function of the expected order statistics. The expectation of an order statistic may be written as

$$EX_{j:r} = \frac{r!}{(j-1)!(r-j)!} \int x[F(x)]^{j-1}[1-F(x)]^{r-j} dF(x).$$

The first few L-moments are

$$\lambda_1 = EX$$

$$\lambda_2 = \frac{1}{2}E(X_{2:2} - X_{1:2})$$

$$\lambda_3 = \frac{1}{3}E(X_{3:3} - 2X_{2:3} + X_{1:3})$$

$$\lambda_4 = \frac{1}{4}E(X_{4:4} - 3X_{3:4} + 3X_{2:4} - X_{1:4}).$$

Note that λ_2 is a measure of the scale or dispersion of the random variable X and λ_3 is the mean difference of the differences between order statistics. It is often convenient to standardize the higher moments λ_r, $r \geq 3$, so that they are independent of the units of measurement of X. Define the L-moment ratios of X to be the quantities

$$\tau_r = \frac{\lambda_r}{\lambda_2}, \quad r=3, 4, \cdots$$

τ_3 and τ_4 may be regarded as measures of skewness and kurtosis respectively. It has shown (Hosking, 1990) that L-moments are more robust than conventional moments to outliers in the data and enable more secure inferences to be made from small samples about an underlying probability distribution.

Since λ_r is a function of the expected order statistics of a sample of size r, it is natural to estimate it by a U-statistic, i.e. the corresponding function of the sample order statistics averaged over all subsamples of size r which can be constructed from the observed sample of size n. Let x_1, x_2, $\cdots$, x_n be the sample and $x_{1:n} \leq x_{2:n} \leq \cdots \leq x_{n:n}$ be the ordered sample, and define the rth sample L-moment to be

$$l_r = \binom{n}{r}^{-1} \sum_{1 \leq i_1} \sum_{< i_2 < \cdots} \cdots \sum_{< i_r \leq n} r^{-1} \sum_{k=0}^{r-1} (-1)^k \binom{r-1}{k} x_{i_{r-k}:n}, \quad r=1, 2, \cdots, n$$

and define the sample L-moment ratio as $t_r = l_r / l_2$, $r=3, 4, \cdots$. Note that t_r is invariant under linear transformation of x_i. The exact sampling distribution of the sample L-moments are difficult to obtain. It can be shown that the asymptotic distribution of $\sqrt{n}(t_3 - \tau_3)$ is normal with mean zero and variance

$$\frac{(\Lambda_{33} - 2\tau_3 \Lambda_{23} + \tau_3^2 \Lambda_{22})}{\lambda_2^2}$$

where

$$\Lambda_{rs} = \iint_{x<y} [P^*_{r-1}(F(x))P^*_{s-1}(F(y)) + P^*_{s-1}(F(x))P^*_{r-1}(F(y))]\ F(x)\ [1-F(y)]\ dx\,dy,$$

and $P^*_r(x)$ being the rth shifted Legendre polynomial such as

$$P^*_r(F) = \sum_{k=0}^{r} p^*_{r,k} F^k \quad \text{and} \quad p^*_{r,k} = (-1)^{r-k}\binom{r}{k}\binom{r+k}{k}$$

3. Small Sample Behavior of the Skewness Measures

The asymptotic sampling distributions of the central moment s_3 and L-moment t_3 are discussed in the previous section. However, those asymptotic sampling distributions are not satisfactory to design a control chart, because the sample size for a control chart is usually small in real situation. The numerical derivations of the exact sampling distributions of s_3 and t_3 are very difficult to obtain. Simulation experiments of the sampling distributions of s_3 and t_3 if the sample size is small are studied in this section. Since s_3 and t_3 are both invariant under linear transformation of a random variable, a standard normal population is assumed in the simulation experiments. Since the sample size of a control chart is usually less than 10 in real applications [Park, Sung Hyun, 1990], sampling distributions of s_3 and t_3 with sample size 4 to 9 are studied. The averages, standard deviations, 0.5 percentiles, 99.5 percentiles and histograms of s_3 and t_3 are reported after 100,000 simulation runs for each sample size.

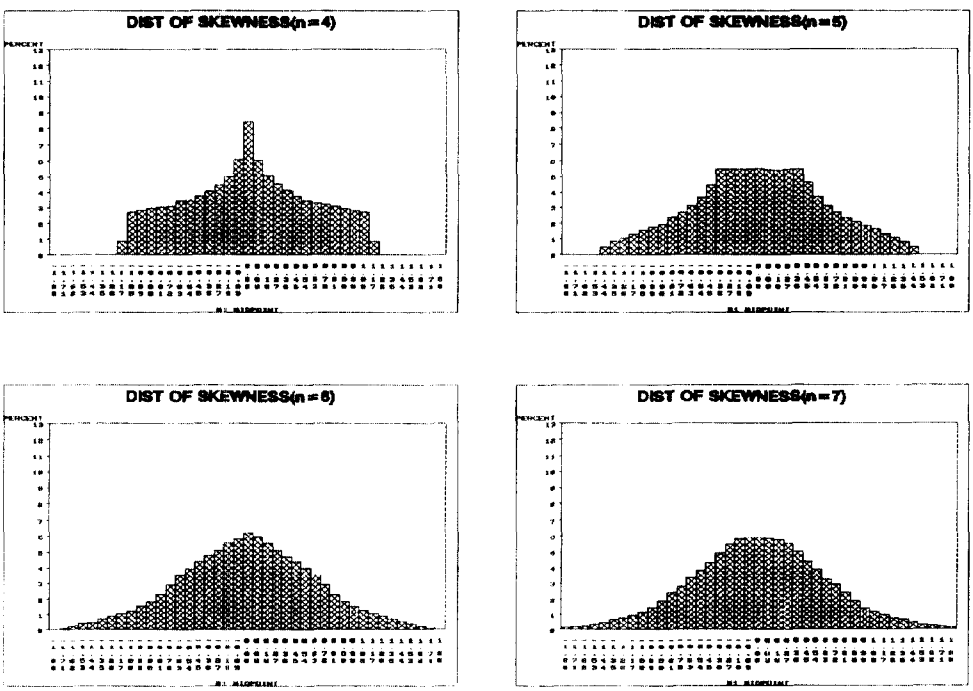
3.1 Sampling Distribution of s_3

〈Table 1〉 shows the statistics of the sampling distribution of s_3 based on

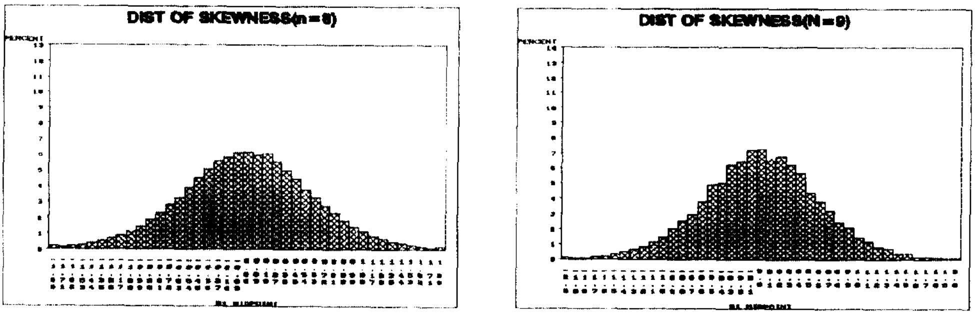
100,000 simulation runs (actually 1000 runs are repeated 100 times) and ⟨Figure 2⟩ shows the histograms for each sample size. The sampling distributions of s_3 are close to normal with mean zero and the standard deviations are getting smaller as sample size increases.

⟨ Table 1 ⟩ The average, standard deviation, 0.5 percentile, 99.5 percentile of the sampling distribution of s_3 for each sample size

Sample Size	s_3			
	Average	Std Deviation	0.5%	99.5%
4	0.001987	0.586074	−1.135629	1.135328
5	0.003333	0.610446	−1.406056	1.383397
6	−0.003334	0.621277	−1.569412	1.528087
7	0.005685	0.617982	−1.604457	1.589196
8	0.006695	0.607809	−1.648214	1.568795
9	0.003235	0.592938	−1.622793	1.592130



⟨ Figure 2 ⟩ Sampling distribution of s_3 for each sample size



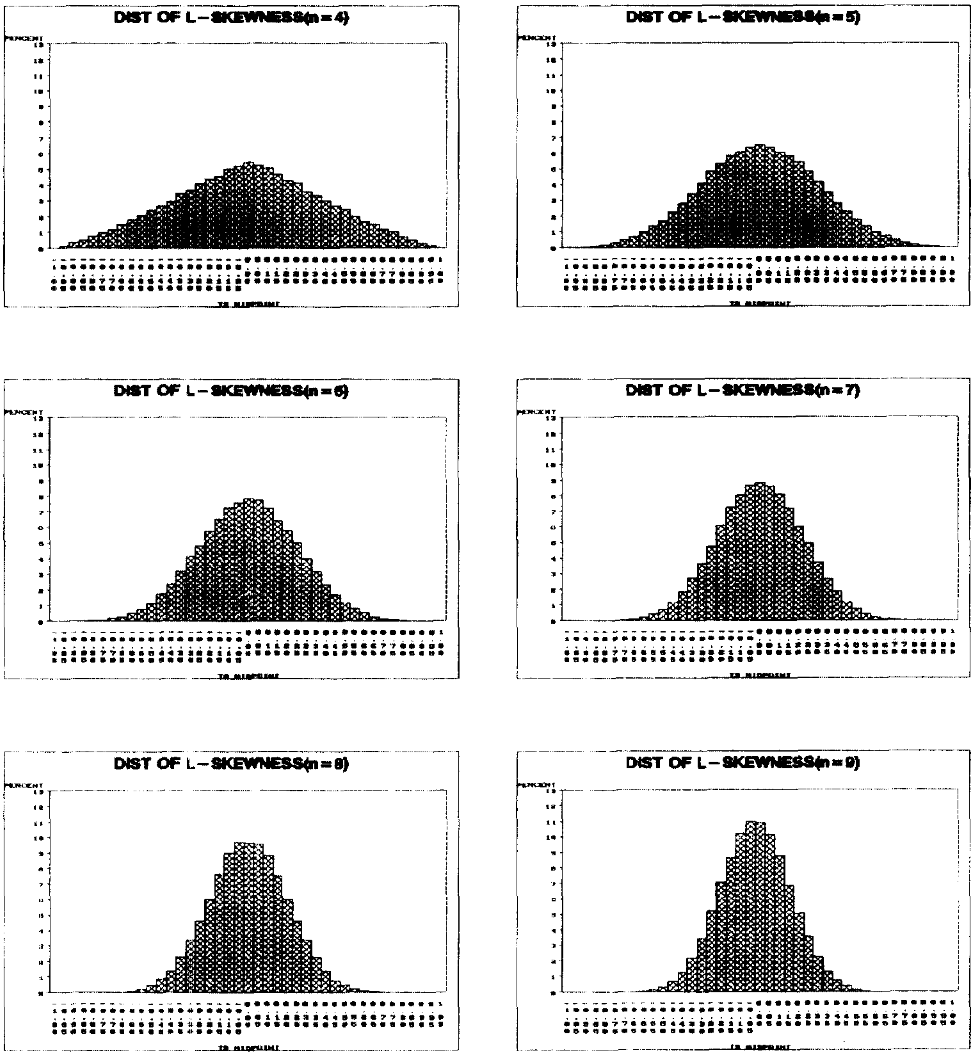
〈 Figure 2-continued 〉 Sampling distribution of s_3 for each sample size

3.2 Sampling Distribution of t_3

〈Table 2〉 shows the statistics of the sampling distribution of t_3 based 100,000 simulation runs (actually 1000 runs are repeated 100 times) and 〈Figure 3〉 shows the histograms for each sample size. The sampling distributions of t_3 are close to normal with mean zero and the variances are getting smaller as sample size increases. However, the standard deviation of t_3 for each sample size is smaller than the one of s_3 .

〈 Table 2 〉 The average, standard deviation, 0.5 percentile, 99.5 percentile of the sampling distribution of t_3 for each sample size

Sample	t_3			
Size	Average	Std Deviation	0.5%	99.5%
4	0.000997	0.376871	−0.869167	0.864059
5	0.001579	0.297747	−0.749898	0.719913
6	−0.001060	0.254997	−0.656390	0.642212
7	0.002269	0.222441	−0.557768	0.571763
8	0.001982	0.200382	−0.515445	0.507029
9	−0.023356	0.179293	−0.482886	0.429459



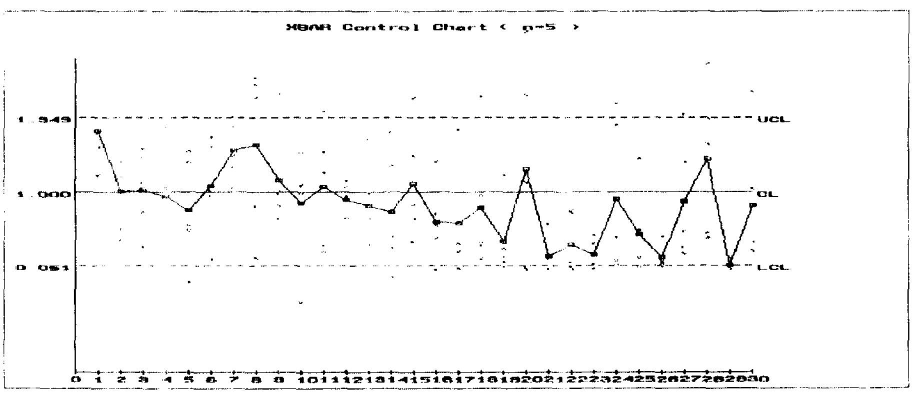
〈 Figure 3 〉 Sampling distribution of t_3 for each sample size

4. Control Chart for Skewness - simulation experiments

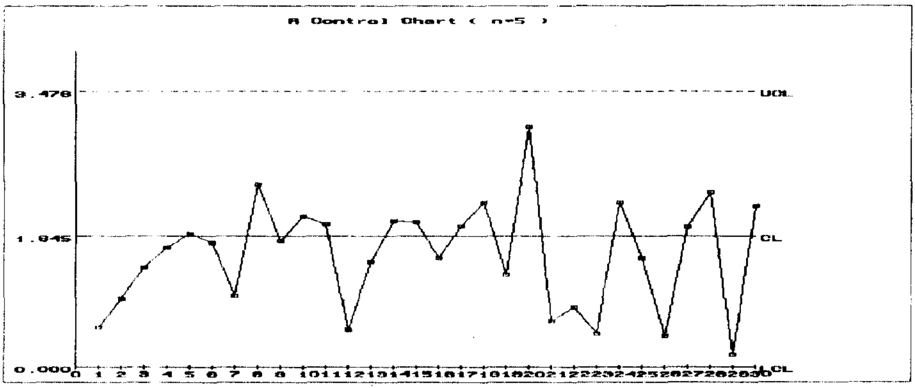
Simulation experiments are done to illustrate the control chart for skewness and to compare the performance of two skewness measures. Assume that the population is normal with mean μ and variance σ^2. Since s_3 and t_3 are invariant under linear transformation, the sampling distributions of s_3 and t_3 in section 3 are used to design a control chart. The 99.5 percentile, average, and 0.5 percentile of the sampling distributions of s_3 and t_3 are used as upper control limit(UCL), center line(CL) and lower control limit(LCL) of the control chart respectively.

Sample data which are under control are generated using the normal random number generator. Sample data which are out of control are generated from the gamma distribution which is skewed to the right. Sample sizes are varied from 4 to 9 and the parameters of gamma distribution are also varied in each simulation experiment. Since we cannot show all simulation results because of page limitations, the case of sample of size 5 and the gamma distribution with mean 1 and variance 0.5 (i.e. $\alpha=2$ and $\beta=0.5$) is illustrated in this paper. This gamma distribution is selected because sample data generated from the distribution seem to be under control based on $\overline{X}$-chart and R-chart. In order to compare with the normal population, the first 15 samples are collected from normal distribution with mean 1 and variance 0.5. The next 15 samples which represent data from out-of control process are collected from gamma distribution with $\alpha=2$ and $\beta=0.5$ which has the same mean and variance.

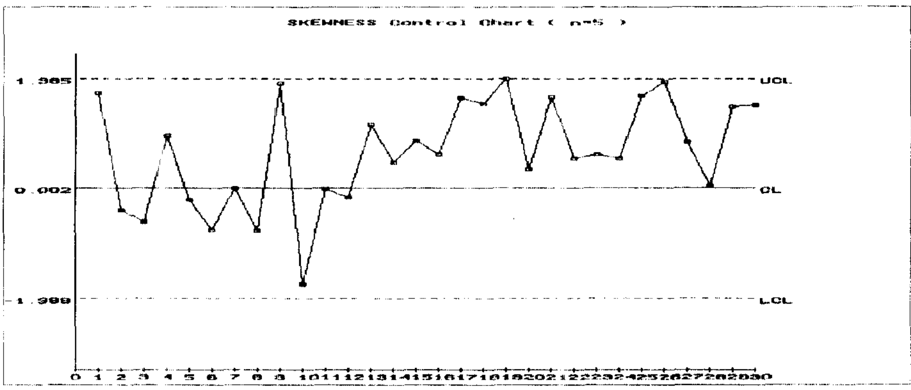
The process seems under control based on $\overline{X}$-chart ⟨Figure 4⟩ and R-chart ⟨Figure 5⟩. However, both s_3 and t_3 skewness charts show that all skewnesses of data from the out-of-control process fall above the center line and some of them fall outside the upper control limit. It means that the skewness control charts would successfully detect the out-of-control process which would not be detected by the $\overline{X}$-chart and R-chart.



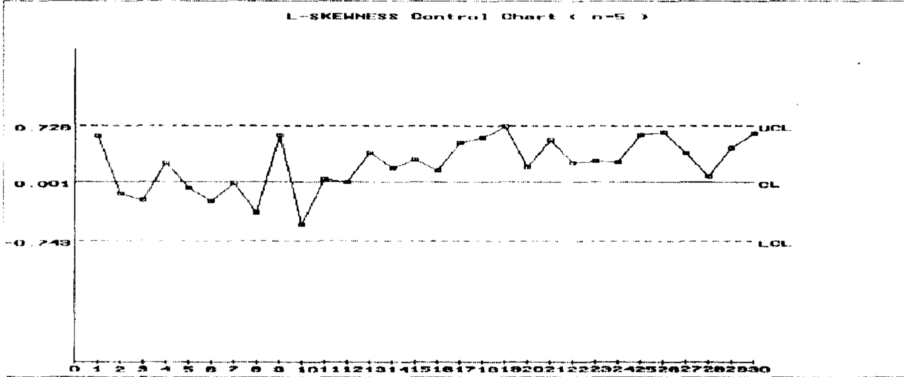
⟨ Figure 4 ⟩ The observed data and $\overline{X}$-chart of the simulated data ---
The first 15 samples are from normal population and the next 15 samples are from gamma population. Both populations have mean 1 and variance 0.5.



〈 Figure 5 〉 R-chart of the simulated data



〈 Figure 6 〉 s_3 skewness chart of the simulated data



〈 Figure 7 〉 t_3 skewness chart of the simulated data

In order to compare the performance of two skewness measures, sample data are generated 1,000 times from the same gamma population($\alpha=2$ and $\beta=0.5$) and the number of samples whose skewness fall outside of the control limit are counted (Table 3). It shows that t_3 skewness chart seems consistently effective to detect outliers than s_3 skewness chart for each sample size. Also, the detection rate of t_3 skewness chart increases rapidly if the sample size increases.

〈 Table 3 〉 Comparison of s_3 and t_3 skewness charts

* samples are generated 1000 times from gamma α=2 and β=0.5

Sample Size	Number of samples fall outside the control limit	
	s_3 control chart	t_3 control chart
4	85	91
5	134	149
6	134	197
7	160	255
8	182	316
9	211	473

Similar experiments have been done for various gamma populations, and the results of the experiments are almost the same as 〈Table 3〉. Hence, we may conclude, at least, that t_3 skewness chart is more effective than s_3 skewness chart to detect outliers in case of gamma population. Although we cannot conclude the same result in other populations at this point, L-moments are known to be more robust than conventional moments to outliers (Hosking, 1990).

5. Conclusion

Empirical sampling distributions of two skewness measures, the third central moment and the third L-moment, are studied using simulation experiments. These sampling distributions are used to construct control charts of skewness. Simulation experiments show the possible use of skewness chart, in addition to $\overline{X}$-chart and R-chart, to detect samples from an out of control process. The third L-moment seems more effective to detect outlier than the third central moment.

All experiments in this paper are based on simulation. Actual performance of the control chart for sample skewness should be evaluated by end users in real manufacturing process. The effectiveness of skewness chart, however, may be understood 'better than nothing' method to check whether a process is under control.

References

[1] Duncan, A. J. (1974), *Quality Control and Industrial Statistics*, Richard D. Irwin, Inc.

[2] Hosking, J. R. M. (1990), "L-moments: Analysis and Estimation of Distributions using Linear Combinations of Order Statistics," *Journal of Royal Statistical Society B*, Vol 52, No 1, pp 105-124.

[3] Park, Sung Hyun (1990), *Statistical Quality Control*, Min Young Sa.

[4] Wilks, S. S. (1962), *Mathematical Statistics*, John Wiley & Sons, Inc.